

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 19, 2000, 23:39:15 ; Search time 4256.96 Seconds
(without alignments)
1508.199 Million cell updates/sec

Title: us-09-339-159-1

Perfect score: 1470
Sequence: 1 ttgaataatggttttaaaaa.....ttgataatgtgattgtaga 1470

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sy:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_htg1:*
31: gb_htg1:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pl4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

BEST AVAILABLE COPY

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pl5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v11:*
73: gb_v12:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pl6:*
92: gb_pl7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|---------------|-------|----------------|--------|----|------------|-----------------------------|
| 1 | 460.8 | 31.3 | 1551 | 1 | AB007123 | AB007123 Escherichia coli |
| 2 | 242 | 16.5 | 1530 | 2 | DB6329 | DB6329 Vibrio cholerae |
| 3 | 134 | 9.1 | 4567 | 1 | AF163837 | AF163837 Caldococcus |
| 4 | 131.8 | 9.0 | 2155 | 2 | CDCMANABD | M36063 C.saccharosus |
| 5 | 131.8 | 9.0 | 4977 | 2 | CDCMANANA | L01257 Caldococcus |
| 6 | 125 | 8.5 | 5284 | 2 | CSU16308 | U16308 Caldococcus |
| 7 | 109.4 | 7.4 | 1461 | 5 | E02075 | E02075 Genomic DNA |
| 8 | 109.4 | 7.4 | 1939 | 2 | BACMANN | M31797 Bacillus sp |
| 9 | 94.4 | 6.4 | 25970 | 2 | SC2H4 | AL031514 Streptococcus |
| 10 | 89.2 | 6.1 | 1881 | 74 | STMBMANASE | M92297 Streptococcus |
| 11 | 81.6 | 5.6 | 837 | 74 | TFU6227 | AJ006227 Thermoanaerobacter |
| 12 | 52.4 | 3.6 | 14922 | 2 | SCF73 | AL121746 Streptococcus |

| | | | | | | | |
|---|----|------|-----|--------|----|-----------|----------------------|
| C | 13 | 48.6 | 3.3 | 7218 | 5 | I66494 | I66494 Sequence 14 |
| C | 14 | 45.8 | 3.1 | 111071 | 53 | AC024591 | AC024591 Homo sapi |
| C | 15 | 45.2 | 3.1 | 35049 | 33 | CELRK3E6 | US5375 Caenorhabdit |
| C | 16 | 45 | 3.1 | 175300 | 40 | AC014459 | AC014459 Homo sapi |
| C | 17 | 44.6 | 3.0 | 4807 | 7 | AB006362 | AB006362 Candida a |
| C | 18 | 44.6 | 3.0 | 143092 | 56 | AC015641 | AC015641 Homo sapi |
| C | 19 | 44 | 3.0 | 14235 | 73 | H2U67264 | U57264 Helicoverpa |
| C | 20 | 44 | 3.0 | 147123 | 54 | AC027030 | AC027030 Homo sap |
| C | 21 | 43.2 | 2.9 | 2138 | 8 | CNT15601 | U15801 Candida alb |
| C | 22 | 42.8 | 2.9 | 148667 | 38 | HS1018E9 | AL035455 Human DNM |
| C | 23 | 42.8 | 2.9 | 251206 | 68 | AL354834 | AL354834 Homo sapi |
| C | 24 | 42.4 | 2.9 | 1063 | 72 | HAU67261 | U67261 Helicoverpe |
| C | 25 | 41.8 | 2.8 | 42539 | 66 | AC078830 | AC078830 Staphylococ |
| C | 26 | 41.8 | 2.8 | 47963 | 60 | AC074320 | AC074320 Staphylococ |
| C | 27 | 41.8 | 2.8 | 54249 | 56 | AC078934 | AC078934 Staphylococ |
| C | 28 | 41.8 | 2.8 | 55418 | 53 | AC025592 | AC025592 Staphylococ |
| C | 29 | 41.8 | 2.8 | 214412 | 43 | AC018695 | AC018695 Homo sapi |
| C | 30 | 41.6 | 2.8 | 17660 | 48 | SC9773 | 248951 S. cerevisiis |
| C | 31 | 41.6 | 2.8 | 38779 | 48 | SCU31900 | U31900 Saccharomyces |
| C | 32 | 41.6 | 2.8 | 107642 | 12 | AF125313 | AF125313 Mus muscu |
| C | 33 | 41.6 | 2.8 | 165536 | 48 | SCCHXVI | Z1255 S. cerevisiis |
| C | 34 | 41.4 | 2.8 | 3734 | 8 | ATRNABE1 | X97970 A. thaliana |
| C | 35 | 41.4 | 2.8 | 81493 | 8 | ATBKIRGEN | X98130 A. thaliana |
| C | 36 | 41.4 | 2.8 | 82646 | 7 | AB028611 | AB028611 Arabidops |
| C | 37 | 41.4 | 2.8 | 183593 | 39 | AC011209 | AC011209 Homo sapi |
| C | 38 | 40.8 | 2.8 | 1302 | 11 | AF090946 | AF090946 Homo sapi |
| C | 39 | 40.8 | 2.8 | 55804 | 49 | HS77N19 | Z98866 Human DNA s |
| C | 40 | 40.8 | 2.8 | 157740 | 53 | AC025003 | AC025003 Homo sapi |
| C | 41 | 40.8 | 2.8 | 191126 | 41 | AC013444 | AC013444 Homo sapi |
| C | 42 | 40.6 | 2.8 | 3804 | 48 | YSCMTCG04 | Y36888 Saccharomyce |
| C | 43 | 40.6 | 2.8 | 3804 | 48 | YSCMTCG04 | J01485 Yeast (S.cer |
| C | 44 | 40.2 | 2.7 | 2103 | 57 | TMU49012 | U49024 Tegeticia |
| C | 45 | 40.2 | 2.7 | 44101 | 60 | AC073960 | AC073960 Homo sapi |

ALIGNMENTS

```

RESULT 1
AB007123 LOCUS AB007123 1551 bp DNA BCT 28-APR-1998
DEFINITION Bacillus circulans gene for mannanase, complete cds.
ACCESSION AB007123
VERSION AB007123.1 GI:3090433
KEYWORDS guar gum-degrading enzyme; mannanase.
SOURCE Bacillus circulans DNA.
ORGANISM Bacillus circulans
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Bacillus.
1 (bases 1 to 1551)
Yoshida,S.
Direct Submisslon
Submitted (08-SEP-1997) to the DDBJ/EMBL/Genbank databases. Seq1
Yoshida, Research Institute of Technology, Konkole construction,
Bioengineering; 4-3-55 Dempo, Konohana-ku, Osaka, Osaka 554, Japan
(tel:06-461-0262, Fax:06-468-3659)
2 (sites)
Yoshida,S., Sako,Y. and Uchida,A.
Cloning, sequence analysis, and expression in Escherichia coli of a
gene coding for an enzyme from Bacillus circulans K-1 that degrades
guar gum
Biosci. Biotechnol. Biochem. 62 (3), 514-520 (1998)
98333274

JOURNAL MEDLINE
FEATURES
Source
1.1551 location/qualifiers
/organism="Bacillus circulans"
/db_xref="taxon:1397"
1.1551
/function="guar gum-degrading enzyme"
/codon_start=1
/transl_table=1
/product="mannanase"
/protein_id="BAA25878.1"

```

| BASE COUNT | 461 a | 333 c | 363 g | 394 t | ORIGIN |
|-----------------------|----------------------------------------------------------------|--------------------|------------|--------------|--------------------------------------------------------------|
| Query Match | 31.3%; | Score 460.8; | DB 1; | Length 1551; | |
| Best Local Similarity | 58.9%; | Pred. No. 1.9e-97; | | | |
| Matches 859; | Conservative 0; | Mismatches 527; | Indels 72; | Gaps 1; | |
| 85 | aacgctaattgcaaatccgcgatttatgttaagcgttaccactcttaagatgccaatgga | 144 | | | "/db_xref="GI:3090434" |
| 91 | AAAGCAATGCTGCGAAGCGGATTATATGTAAGCGGTACCAATTATGAGTGTCTACAGA | 150 | | | "/translation="MGWFLVTLRRWLIATFAFLMFESWTGQLINKAHAAAGCFYWGCTT |
| 145 | aaccatttgtaattagaagggaattaacatagggcgcgatgataaagaccgaagcaat | 204 | | | LIDATGQFVNRGVNHAHTWYKRDLSIAPIAATGANTITIVLANGKRWLDVNTW |
| 151 | CAACCATTTGTGTATCGAGAGATCAATCATGGCACATGSTATAAAGATCAACTATCC | 210 | | | NNILITGQNKLIIVLEVHADTSGDSLDLNAWYMGKISALIGEDRYIIMANE |
| 205 | actgcaattggaagggaattgcaaatccggtgtctaatacgcgtccggattggttatcgtat | 264 | | | WYMGWGAAMAGAKQAIPIKRNAGLFTLLTVDSAGAMQOYDSKNGTETVNDLPKE |
| 211 | ACCGCAATACCAAGCATTGTCTAAACAGGTGCCAACAGATACGATATGTTACTGGCGAT | 270 | | | NTVSIIMHYEAXGANDSVKNSINDGVINKNLALIIGEGGNGHTKDEATIMTSOE |
| 265 | gggggacaattgacaaaagatgacatccatcacagtaagaagaacctatctcttagcgga | 324 | | | KGVGMASWKSNGNSDLAYIDMTMDMAGNSLITSGONTVYNSNOKIKATSVLSGIFGCT |
| 271 | GGACACAATATGACGCTTGATGATGCTTAACACGCCGCAACATATTTCTACCCCTCTGTAA | 330 | | | TPISSTPTSTPTSTPTSTPTSTPTSPSPSPSPSNGNGITLLYDETGTGMSGNISGEPWAT |
| 325 | gataatcattggtgtgctgtcttcgttaagttcaatgatagtctacccggttatcattcgtc | 384 | | | EMKATQATTLKADVSLQSNSTHSLYITSNOLSKRSLSIAKVAIKMANGNITNGIYAKH |
| 331 | CAAAAACAACATTAATGTCGGTTTGGAACTACATGACGCTTACAGGAACGATAGTCTTTC | 390 | | | YVKIIGSGTWYDSEMLTQSNIDGTLILTLISGINSLSVKIEIGVEFRASSNSQSAIA |
| 385 | tcgcctcaatcgctgctgtgattatgattagaattagaagatgcttattatggaagaa | 444 | | | YVDSVSLQ" |
| 391 | GATTTAGACACAGCCGCTTAATTACTGCGATTGCTTTAAACGCGGTTGATCGCAGGAA | 450 | | | |
| 445 | gataccgcatlataatattcgcaatgaatgattggttctgtyggaagggaagtcttg | 504 | | | |
| 451 | GACCGTGAATCATTAATATATAGCTAACAGAGGTATACGAAACATGGAGTGGATGCGCTGG | 510 | | | |
| 505 | gcctgacgggtttaaacaagaacatccgcgattgtgtaacgcggtctaaacaactctg | 564 | | | |
| 511 | GCATATGCTTATTAAGCAACGCTATACCCAAATCGGTATGATGCTTAACTCATACAGCTG | 570 | | | |
| 555 | atggtaagatgctcgggggtggaacaattccacaatcogattcattatgataagaa | 624 | | | |
| 571 | ATTGTTGATCCGCTGGATGGGAGCAATATCCGAATTCGGTCAAAATTTATGGAGCAAA | 630 | | | |
| 625 | gttttaattgctgacccctcaagaataacaaatgttttcogattcatglatgtaattgca | 690 | | | |
| 631 | GTACTGAATGACAGCCGCTTAAAAAACACAGTATTTCTCTAACCATATGATGATTAACCT | 690 | | | |
| 685 | ggtgtaatgcatcgcaaatctcgtaactaataatgacgaagttcttaataagacctgca | 744 | | | |
| 691 | GGGGGCATGCAATATACCGTCAATATTCATATTTACGCTGTGCTGAACACAAATCTTGCA | 750 | | | |
| 745 | ttagtcatgttgtaatttgacacccgtcatatacaaatggtgacgtcgtatgaagcaagat | 804 | | | |
| 751 | CTGATTTATCGCGCAATTTGGTGGACAACATATCAACACGCTGATGGATGAAGCCACCAT | 810 | | | |
| 805 | atgagcattctgacaagaagaggtgggtggtgctgctggtcatatgaaaggaaagcagc | 864 | | | |
| 811 | ATGAGTTATTTCCCAAGAGAAAGGAGTGGCGCTGGTGGCTTGCTTGAAAGGCAATAAGC | 870 | | | |
| 865 | ccaagatggagatatttagaccttcgaatgattgggctgtagaaataaccttaacgcttg | 924 | | | |


```

Db      871 AGTATTTGGCTTATCTCGATATGACAAATGATGGCGTGTACTCCCTCAGCTTC 930
Oy      925 ggaataacataatgtaatggtccataatggttaagaagaactcgagattgaacccgtt 984
Db      931 GGTAAATACCGTAGTAAATGGCGATGTAACGGCATTAAGACACTTGTGTATCCGGCAT 990
Oy      985 tttaacaggtagagatctgtagtgaagactctcgcaactc----- 1027
Db      991 TTGGAGGTTGTTACGCCAACCCTCAAGCCCTACTCTACACCTACATCAACGCCAACCTCA 1050
Oy      1028 -----tttat 1032
Db      1051 ACTCTACTCTCAGCGCAAGTCCGACCCGAGTCCAGGTAATACGGGACGATCTTATAT 1110
Oy      1033 gattttgaagtagtatgcaaggaatggaactggaagtagcttgagcggaggtcttggtc 1092
Db      1111 GATTTCGAAAGAGCAAGCAAGGCTGTGCGGAAACAATATTTCCGGAGGCCCATGGGTC 1170
Oy      1093 gtagacaggtggtctctcaagaagaagtcattctttaaagaaggatccaattgctgc 1152
Db      1171 ACCAATGATGAAAGCAACGCGGACGCAAACTCTCAAGCCGATGCTCTTACAAATCC 1230
Oy      1153 aattcacacattactacatgltatccaataacgctctttacagcagaatagtagata 1212
Db      1231 AATTCACAGCATAGTCTATATATATACCTCTAATAAATCTGTGCGAAAAAGAGTGTG 1290
Oy      1213 caagctactgttaaacatgcaaatgagggaagtggttgtaagtgaatgactgagcgtctt 1272
Db      1291 AAGGCAACGGTTAAACATACGCAACGCGGCAATATCGGCAACGGGATTTATGCAAACTA 1350
Oy      1273 tatgtgaagaagacatggttatatactgtaactggaagcttgtagcgaatagcgt 1332
Db      1331 TACGTAAGACCGGCTCCGGGTGACATGTAAGATCCGGAGACAATCTGATTCATCA 1410
Oy      1333 tcatctggaacacacgctatctctagattatccaatgccaatactcttccaagtaag 1392
Db      1411 AACGACGATACCATTTTACACTATCCCTCAGCGCATTTGCAATTTGCCCATCAAA 1470
Oy      1333 gaaatgtagtccatcccaatcagcaggtgtagtagtaggaacaaactcgattatctt 1452
Db      1471 GAAATTTGGGTAGATTCGGCCCTCTCAACAGTACGAGGCAATCTTATATGTA 1530
Oy      1453 gataatgtagttagaa 1470
Db      1531 GATAGTGTACTCTGCA 1548

RESULT 2
D86329 1530 bp DNA BCT 19-MAR-1998
LOCUS D86329 1530 bp DNA BCT 19-MAR-1998
DEFINITION Vibrio sp. gene for beta-1,4-mannanase, complete cds.
ACCESSION D86329.1 GI:2978250
VERSION beta-1,4-mannanase; manA.
KEYWORDS
SOURCE Vibrio sp. (strain:MA-138) DNA.
ORGANISM Vibrio sp.; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 1530)
AUTHORS Tamaru, Y.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1996) to the DDBJ/EMBL/GenBank databases. Yutaka
Tamaru, Mie University, Faculty of Biosciences, Utilization of
Marine Products, 1515 Kamihama, Tsu, Mie 514, Japan
(E-mail: s9tamaye@maric.dio.mie-u.ac.jp, Tel: 0592-31-9561,
Fax: 0592-31-9557)
2 (sites)
REFERENCE Tamaru, Y., Araki, T., Morishita, T., Kimura, T., Sakita, K. and
AUTHORS Ohnishi, K.
TITLE Cloning, DNA sequencing, and expression of the beta-1,4-mannanase
JOURNAL gene from a marine bacterium, Vibrio sp. strain MA-138
FEATURES
location/Qualifiers
source 1..1530

```

```

/organism="Vibrio sp."
/strain="MA-138"
/db_xref="taxon:678"
-35_signal
-10_signal
RBS
gene
CDS
136..141
149..1336
/gene="manA"
149..1336
/gene="manA"
/codon_start=1
/transl_table=1
/product="beta-1,4-mannanase"
/protein_id="P42518.1"
/db_xref="GI:2978251"
/translation="MKETKAIISLLEITMASCAHAGFYVSGVLYEANGSAFKIRGIN
HAHWYDQKLSVALSGIAAGANTRVVLSNGYRMTKNDVSDVINILNKKANLILAI
LEVHDTTGGESSAASLDSADADYIELKRELIGEDYVILINLEPFGNNDAVAT
NDHSAIORLSRAGINHTIWDAPNMGDMKFWLNAOFVFNDSPLKLTIFSTYV
VYSYNSVNDYISSFTNNGLYLVYGERASTHKGADNDGSIKMSERTLSGYTGRS
GNDTTSDDLIDVNMWNNSTSTMGVNLINGONGIKSTLTATVTCNDKNDPDSCTV
PICSSAVDPDGDGWMENNOCIVDSSDTPAPNGYPCSQESSDPDGDGWMENNNS
MCC"
BASE COUNT 469 a 211 c 333 g 517 t
ORIGIN
Query Match 16.5% Score 242; DB 2; Length 1530;
Best Local Similarity 57.2% Pred. No. 2.4e-46;
Matches 525; Conservative 0; Mismatches 375; Indels 18; Gaps
Oy 93 tgcgaatccggattttatgtaagcgggtacacactctaacgtagtgcgaatggaacccctt 152
Db 202 TGCTCATCGGGGTTTATGATCAAAATGCTCTTATGAAAGCAATGATGCTT 261
Oy 153 tgaatgagaaggaatcaaccatgagcagcatgataaagaccagcaactactgcaac 212
Db 262 TAAATTTAGAGGATTAATTCATGCATCTTGATGATACAGCAAGTTAAACGTGGCTT 321
Oy 213 tgaaggatctcaaatcagcgtgtaatacagtcgccgtagtgcgtatctgtagtgggaca 272
Db 322 AAGTGTATGCTGCCACAGGACCAATCTGTACGGGTTGACTAAGTAATGATTATCG 381
Oy 273 atgacaaaagaatgacatacagtaagaacctatctcttaaggagaataatca 332
Db 382 GTGCAAAAATATGACGTTAGATGATGATCACTTAATTAATTAAGCAAAAGCAAA 441
Oy 333 ttgtgtcgtctctgaatgtagtgcacgggtatg-----atccattgc 383
Db 442 CCTTATACCTATATTAGAAATTCATGACACTACCTGATGAGAGAAAGATAGTGTGC 501
Oy 384 ttgcgtcaatcgtcgtctgtagtattgtagtgaatgagaagtgcttcaattggaagga 443
Db 502 TAGTTAGATCTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 561
Oy 444 agatacgcattatttaataatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 500
Db 562 AGATATGTCATCATTAACCTACGCAATGAACCTTTTGTGATTAATTAATGATGCGTGC 621
Oy 501 ttggctcagcggatataaacaacaccccgcatgtagtaacgcggcttcaaacatcac 560
Db 622 TTGGTTAATGACCATGTCAGTGCATTCATCAACGATTAAGAAAGTGCAGGATTAACAC 681
Oy 561 ctgtagttagatgctgctgggggtggaacaa---ttccaatcgtatcattgattag 617
Db 682 GATTATGTTGATGCTCCAAACTGGGGGCAAGACTGGAAGGGCTTTATGCTGAATATATGC 741
Oy 618 aagagaagtttttaattgctgacccctcaacgaaatacaatgttttcgattcatatgtaga 677
Db 742 TCAATTCGTTTAAATTTGACCCCTTAATTAATGAATGATTTTGTGTTCAATGATGAGA 801
Oy 678 atagcaggtgtagtgcgaaggtgcacaaatgtagcgaaggttcgaatgtaga 737

```



```

QY 218 ggatgcaataaccggtgctataagcgttcggaattgtgtatctgtatggggaatgga 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2130 GCATTGCTGCTGGGGCCGACCGCGTGGCTGCTGACGACGCGTGGCGGTTG 2189
QY 278 caaagaatgacatccatgaagaacctatctctttagcggagaataa---catt 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2190 CGAAGATTCCGGCCACGACGAGTACGACATTTTCTCAGCGCGGACGCTCGTTATC 2249
QY 335 tgggtcgtctcttgaagtcacatgacgttaccggtatgattcattgc-----tt 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2250 GGGCGGCTGCTTGAAGTTCACGACACGCGGTACGCGGGAAGACGCGCGCGTGT 2309
QY 386 cgtcaatcgtctgtgtgtattgtatgaaatgaaagtgtcttaattgaaagaag 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2310 CGATGACGACGGCGGCTCAATTATGATTGAAGAACGCTGCGCGGTACGAGAA 2369
QY 446 ataccgtcatataatattgcaaatgtaattgtgttcgttggaagggaatgctgg 505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2370 ATTTCGTCTGCTCAATATCGCAATGACCGTACGCGGAACTAATTTATCAGAACTGG 2429
QY 506 ctgaacggtataaacaagaatcccgatctggtacacgctcctaaacatcctga 565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2430 TCAGGACACGCGGACGCGGTCCAGCGCTGCGGGAACGCGGGATCAACAATACATCA 2489
QY 566 tggtagatctcgtgggtggaac---aattccacatcgatcattatgaaag 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2490 TGGTGACACCTCGGAACCTGGGTGACGAGCTGCTTCAACGACGCGGACGCGCGCA 2549
QY 623 aagtttatactgacccctcaacgaataatgcttcgattcatatgataatg 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2550 CGATTTCACGCGGACCGCGGACGACGACCTGCTTTTCATTCACATGTACGGG---G 2606
QY 683 caggtgtaatgacatcgcaatctgactaataatgacgaagcttcaatcaagacctg 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2607 TGTACACACGCGGCGGAGAGTCAAAATTAATGAGTCGTTCAAAACCGCGGTGCG 2666
QY 743 catagcatctgtgaattggaacacgctcaataaattgtgacgtcgatgaagcaag 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2667 CCTGTGTATCGGGGAATTGTGCATATGACACGCGCGGTGCGGAACGACGCGCA 2726
QY 803 tcatgagcatctctgacaagaagaggtggtgtgtgctggtacgaaagaag 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2727 TCGTACAAATATCGAAGCAATACACATTGCTGTTCGCGTGTGTGACGCGGTAACG 2786
QY 863 gcccaaatggaatgataccttcgaatgattggtggtgaaataccttacagctt 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2787 GAGCGGCGCTGAATATTGACATGTGTACGACGATTCATATGACGACCGACGCGGT 2846
QY 923 ggggaa 928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2847 GGGGGA 2852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
LOCUS CDCMANABD 2155 bp DNA BCT 26-APR-1993
DEFINITION C.saccharolyticum beta-D-mannanase (mana) gene, complete cds.
ACCESSION M36063.1
VERSION M36063.1 GI:144292
KEYWORDS 1.4-beta-D-mannan mannanohydrolase; beta-D-mannanase.
SOURCE C.saccharolyticum DNA, clone lambda-NZP2.
ORGANISM C.saccharolyticus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Caldicellulosiruptor saccharolyticus
Caldicellulosiruptor group; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 2155)
AUTHORS Luehl, E., Jasmat, N.B., Grayling, R.A., Love, D.R. and Bergquist, P.L.
TITLE Cloning, sequence analysis, and expression in Escherichia coli of a
gene coding for a beta-mannanase from the extremely thermophilic
bacterium Caldocellum saccharolyticum
JOURNAL Appl. Environ. Microbiol. 57, 694-700 (1991)
MEDLINE 91247819
COMMENT Draft entry and computer-readable sequence kindly submitted by

```

```

E Luehl, 03-JUL-1990.
Author address: E Luehl
University of Auckland
Department of Cellular and Molecular Biology
Private Bag
Auckland
NEW ZEALAND
email: SYSTEM@CELLVAX.AUKUNI.AC.NZ.

FEATURES
source
1..2155
/organism="Caldicellulosiruptor saccharolyticus"
/db_xref="taxon:44001"
<1..685
/codon_start=2
/label="ORF1"
/protein_id="AA072860.1"
/db_xref="GI:144293"
/translation="GSNTWRFQMSMRVAEYVYVGDKNAGALLERKWSMISYVX
LNSDSTRIPSTLDSGPDPTWNGTITGPNLHVAVVDYDGLDITLALNLLYISA
GTRKYGFDEERKNLAKELDRMKLYRDEGLSAPERKADYKRFEEQVYIPAGWTG
KMPNDVIRKSGVKFIDIRSKYKODPDMPLBAAYKSGVPERFRRHFWAACDIALVNA
TYELIFGNO"
RBS
818..830
/label="830"
/feature="830"
/feature="mana ribosome binding site"
sig_peptide
841..963
/product="beta-D-mannanase"
CDS
841..1881
/EC_number="3.2.1.78"
/codon_start=1
/transl_table=1
/product="beta-D-mannanase"
/protein_id="AA072861.1"
/db_xref="GI:144294"
/translation="MRKTKIRKMLSVLCYVFLNLTILFANTVILPKVGAATSDNG
VYKIDTSLTILNHAHCWYRDLRDLALGIRSMGNSVRLVNSGRYRKLPAEAVN
IISLSRSGFRAIILEVHDTTGYGSDGACISLAQVEYKELKSLDNEPVIINIG
NEPYNNTQWVNDTNAIKALRDAGKHTIWDAPWMDGWSMTMDKNSIHEAD
PLRLNLFVSIHMYGVNTASKVEEYIKSFVDGLPLVIGFQHDTGDPDEAIVRA
KOYKIGLFSWMSGNSYVGLDWNMNDPNNPFPWGMVKYTNALIGTSSTPTSTVT
PMPPRHQRHQR"
mat_peptide
964..1878
/product="beta-D-mannanase"
BASE COUNT 721 a 334 c 590 g 510 t
ORIGIN
Query Match 9.0%; Score 131.8; DB 2; Length 2155;
Best Local Similarity 51.7%; Pred. No. 1.3e-20;
Matches 407; Conservative 0; Mismatches 362; Indels 18; Gaps 4;

QY 158 tgaaggagattacacatgagcagcagatgatacaagaacgaactactgcaattgaag 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 998 TAATAGGACCAATACGACATTCGCTGTACAGAGTACTGTATACGCGATTGCCG 1057
QY 218 ggaatgcaataaccggtgctataagcgttcggaattgtgtatctgtatggggaatgga 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1058 GAATTAGGTCATGAGGATGATGAACCTGTGAGGGTGTGCTGATGATGCTTACCGATGGA 1117
QY 278 caaagaatgacatccatgaagaacctatctctttagcggagaataa---catt 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1118 CGAAGATTCCGACGACGAGTACGACATTTTCTCAGCGCGGACGCTCGTTATC 1177
QY 335 tgggtcgtctcttgaagtcacatgacgttaccggtatgattcattgc-----ctt 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1178 AAGCTATTATTAGAAGTCACGACACAGATGATGAGAAAGTGGCGACGATGTT 1237
QY 386 cgtcaatcgtctgtgtgtattgtatgaaatgaaagtgtcttaattgaaagaag 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1238 CATTTGACACACGAGTGAATTTGGAAGGAGATTAAGGCTATTAGACGCTAACGAA 1297
QY 446 ataccgtcatataatattgcaaatgtaattgtgttcgttggaagggaatgctgg 505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

| | | | |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|------|
| Db | 1298 | ATTTTGTAATTAATAACATTGGTAATGAAGCCGTAATGGCAACAATACTATCAAAACTGGG | 1357 |
| Oy | 506 | ctgacgggtataacaagaacatcccgagttgcgtaacgccgylctaaccactctga | 565 |
| Db | 1358 | TTAATGCACACNAAGAAGCATATAAAGCATTAGAGATGACAGATTTCMAACACACGATAA | 1417 |
| Oy | 566 | tgtatagatcgtcgsgggttggagac---aatlccacatcatgatcatgatattgaaag | 622 |
| Db | 1418 | TGTGTGATGCCCGCACTGGGGTTCAGAGATTGCTTAATACATGAGAGATAATGCCAGA | 1477 |
| Oy | 623 | aagttttaatcgtcacacctcaacgaatacatagtittcgatcataatgatataatg | 682 |
| Db | 1478 | GCAATATGAGAGCAGATCCGCTCGCATTTTGTAATTTTCATTCATATGATATG--GCG | 1534 |
| Oy | 683 | cagtggttaatagtcacgcgaagttcgtactaatatgacccgggtcttaataaacctcg | 742 |
| Db | 1535 | TATACAAATACAGCAAGCATCGAAGATCATCATCAATCATCTTTGTTGATTAAGGGGTTAC | 1594 |
| Oy | 743 | cattaagtcattgvtgaatttgagacacgclcatacaatlytgaactgcataagaacaga | 802 |
| Db | 1595 | CATGGTATTATGGGAATTTGGACATACGACACAGATGGTAGCCCGATGATGAAGACTTA | 1654 |
| Oy | 803 | ttatagactattctgacaagaaggagttggttggttggtgctgtgcatatgaaagaaag | 862 |
| Db | 1655 | TTTGACAGTATGACCAAAACAGTACAAAGATAGATTAATTTAGTTGGTCGGTGGTGAATT | 1714 |
| Oy | 863 | gccacgaatgggagatttagaaccttcogaatgatitggcgtggaataacctcaagctt | 922 |
| Db | 1715 | CGAGCTATGTTGGGTATTGGACATGTAATCAACACTGGGACCACCAATATCAACTCAT | 1774 |
| Oy | 923 | ggggaaa 929 | |
| Db | 1775 | GGGGACA 1781 | |
| RESULT | 5 | | |
| CDCCMANA | | | |
| LOCUS | | | |
| DEFINITION | CDCCMANA 4977 bp DNA BCT 26-APR-1993 | | |
| ACCESSION | Caldocellum saccharolyticum beta-mannanase/endoglucanase (mana) | | |
| VERSION | L01257 | | |
| KEYWORDS | gene, complete cds. | | |
| SOURCE | L01257.1 GI:144290 | | |
| ORGANISM | beta-mannanase; endo-1,4 beta-mannanase; endoglucanase; mana gene. Caldocellum saccharolyticum (library: lambda NZP2) DNA. Bacteria; Firmicutes; Bacillus/Clostridium group; Thermosphaerobacter group; Caldiceillobiosruptor. | | |
| REFERENCE | 1 (bases 1 to 4977) Gibbs,M.D., Saul,D.J., Luthi,E. and Bergquist,P.L. The beta-mannanase from Caldocellum saccharolyticum is part of a multidomain enzyme | | |
| AUTHORS | Appl. Environ. Microbiol. 58, 3864-3867 (1992) | | |
| TITLE | Location/Qualifiers | | |
| JOURNAL | 1. 4977 | | |
| MEDLINE | /organism="Caldiceillobiosruptor saccharolyticus" | | |
| FEATURES | /db_xref="taxon:44001" | | |
| source | 841..4836 | | |
| gene | /gene="mana" | | |
| CDs | 841..4836 | | |
| | /gene="mana" | | |
| | /EC_number="3.2.1.78" | | |
| | /note="endoglucanase EC_Number= 3.2.1.4" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="beta mannanase" | | |
| | /protein_id="AAA71887.1" | | |
| | /db_xref="GI:144291" | | |
| | /translation="MRKTKIRKKWLSVLCYVELLNLITLTANTYTLPEKYGAATSDNGC VKRTSTLLIGNNHWCWRDLDTPLALGISWGNSYRVVLISNGYRWTKLPASEEVANA TISRSGAFKAILIEVDHTGYGDGAGASLAQAVEYKEKTSYLQDNEFVLINIGL NPPYRNNTNYQNVNDTNKAIRKLRPGCKRHITMDAPYMGDSMTMADNOISTMEDA PLRNIVESTHMIGVTNASKVEETIKSYVDKGLPIVIEFGHRHDGPDEALYRRYA | | |

[illegible]

Db 1655 TTGTGAGTATGCAAAAGATGATGATTTAGTTGTCGTGCTGTGGAATT 1714
QY 863 gccacgaatggagatttaagaccttcgaatgattggctggaataaccctacagctt 922
Db 1715 CGAGTATGTTGGTATTTGACATGATTAACACTGGGACCCCAATATCAACTCAT 1774
QY 923 ggggaaa 929
Db 1775 GGGGACA 1781
RESULT 6
LOCUS CSU16308 5284 bp DNA BCT 10-FEB-1996
DEFINITION Caldocellulosin pseudogene, complete cds.
ACCESSION U16308
VERSION U16308.1 GI:577827
KEYWORDS
SOURCE Caldocellulosin saccharolyticum.
ORGANISM Caldocellulosin pseudogene, complete cds.
REFERENCE 1 (bases 1 to 5284)
AUTHORS Morris, D.D., Reeves, R.A., Gibbs, M.D., Saul, D.J. and Bergquist, P.L.
TITLE Correction of the beta-mannanase domain of the celC pseudogene from
JOURNAL Appl. Environ. Microbiol. 61 (6), 2262-2269 (1995)
MEDLINE 95314280
AUTHORS Morris, D.D.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1994) Daniel D. Morris, School of Biological
Sciences, University of Auckland, Private Bag, Auckland 92019, New
Zealand
FEATURES
source location/qualifiers
1..5284
/organism="Caldocellulosin pseudogene"
/db_xref="taxon:44001"
misc_feature 279..341
/gene="celC"
/note="encodes leader sequence"
gene 279..4374
/gene="celC"
CDS 279..4374
/gene="celC"
/note="The celC gene potentially encodes a multidomain
enzyme comprised of an N-terminal endoglucanase domain,
two central cellulose-binding domains, and a C-terminal
mannanase domain. However, due to the presence of two
open-reading frame shifts in the mannanase-encoding region
of celC, the actual enzyme is truncated and comprises only
the endoglucanase and cellulose-binding domains"
/pseudo
/product="endoglucanase"
misc_feature 279..3572
/gene="celC"
/note="ORF #1 from celC start codon to frameshift #1"
misc_feature 342..1715
/gene="celC"
/note="encodes endoglucanase domain"
misc_feature 1722..2204
/gene="celC"
/note="encodes cellulose-binding domain (CBD) number 3"
misc_feature 2205..2342
/gene="celC"
/note="encodes proline-threonine linker number 1"
misc_feature 2343..2807
/gene="celC"
/note="encodes cellulose-binding domain (CBD) number 1"
misc_feature 2808..2957
/gene="celC"

misc_feature /note="encodes proline-threonine linker number 2"
2958..3419
/gene="celC"
misc_feature /note="encodes cellulose-binding domain (CBD) number 2"
3420..3524
/gene="celC"
misc_feature /note="encodes partial mannanase domain"
3420..3524
/gene="celC"
misc_feature /note="encodes proline-threonine linker number 3"
3574
/gene="celC"
misc_feature /note="location of the first -1 frameshift"
3575..3625
/gene="celC"
misc_feature /note="ORF #2 from frameshift #1 to frameshift #2"
3627
/gene="celC"
misc_feature /note="location of the second -1 frameshift"
3628..4374
/gene="celC"
BASE COUNT 1829 a 881 c 1238 g 1336 t
ORIGIN
Query Match 8.5%; Score 125; DB 2; Length 5284;
Best Local Similarity 51.1%; Pred. No. 5e-19;
Matches 405; Conservative 0; Mismatches 370; Indels 18; Gaps 4;
QY 211 attgaagagatgcgaataacacggtgctaatatcgccgagatggttctcgaatgggga 270
Db 3610 ATTGGTGGGAATCAGGCGATGGGGATGAACTCTGAGAGGATGCTAGATAGCTTAC 3665
QY 271 caatgacaaagatgacatcatcagtaagaacattcctcttagcggaagataat 330
Db 3670 CGATGACGAAGATACGACGACGATGACGACGACGACGACGACGACGACGACGAC 3729
QY 331 cattggtgctgtt---ctggaagtcagatgacggtgctgctgctgctgctgctgctg 382
Db 3730 GGTTCGAAGCTATATATATATATATATATATATATATATATATATATATATATAT 3789
QY 383 ----cttgctcaatcggtgctgctgctgctgctgctgctgctgctgctgctgctgctg 438
Db 3790 GCGTGTCTATGTAACAGCAGTGAATGGAAGGATGAAGGATGAAGGATGAAGGAT 3835
QY 439 aaggaagatccgctcatcattatcattcgaatgagatggttgcgtggaaggggat 438
Db 3850 AACGAAGATTTTGTATTTATTTAACTTTGCAATGAGCCGTATGGCAATTAATCA 3900
QY 499 gtttggctgacggtgctataaacaagaatcccgatgctgtaacgcccgtctaaacat 554
Db 3910 AACTGAGTTATGACACGACGACGACGACGACGACGACGACGACGACGACGACGAC 3965
QY 559 accttgatgtagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 600
Db 3970 ACGATTAATGGTGAGGCCGCCCACTGAGGAGGATGAGGATGAGGATGAGGATGAT 4020
QY 616 ggaagagaagttttaaagtcgaccccaacgaataacattgcttgcattatgctat 675
Db 4030 GCCAGACATTAATGGAAGCAGATCCCTGCGCAATTTGGTATTTTCCATTCATATGAT 4089
QY 676 gaatacgagtgatgataatgacgcaatgctgctgctgctgctgctgctgctgctg 735
Db 4090 G---GTGATACAAATACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4146
QY 736 gacctcgatgtagtattggtgatttgacacgctcatcaacaagtggaagtgatgaa 795
Db 4147 GGGTACCATTTGTTATTTGGGGAATTTGGACATGACGACGACGACGACGACGACGAC 4206
QY 796 gcaacgattatgagctatcctgcaacaagagaggttgggtggtggtggtggtggtg 855
Db 4207 GAAGCTATTGTCAGGTATGCAAAACGATGCAAGATAGATATTATTTAGTGGTGGTGT 4266

```

OY 856 gggaacggcccaagatgagatattagaccttcgaatgattggcgctggaataacctt 915
DB 4267 GGAATTCGACGATGTTGGGATTGTCATGTAACCACTGGAGCCCAATATATCCA 4326
OY 916 aaagcttgggaataacataagatgattgctcacaatgattgaagaaacttcagatta 975
DB 4327 ACTCATGGGGGCAATGATATTAACATATGCAATGTGTACAAAGTGAACATAATATA 4386
OY 976 agcaccgttttta 988
DB 4387 TTCAAGTGTATA 4399

RESULT 7
E02075 1461 bp DNA PAT 29-SEP-1997
LOCUS E02075 Genomic DNA encoding beta-mannanase of Alkalophilic Bacillus
DEFINITION sp.-AM-001.
ACCESSION E02075.1 GI:2170317
KEYWORDS JP 1989228477-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1461)
AUTHORS Akino,T. and Horikoshi,K.
TITLE BETA-MANNANASE GENE DNA, RECOMBINANT PLASMID CONTAINING SAID DNA AND
TRANSFORMANT
JOURNAL Patent: JP 1989228477-A 1 12-SEP-1989;
RES DEV CORP OF JAPAN, AKINO TOSHIO, HORIKOSHI KOKI
OS Alkalophilic Bacillus sp.
PN JP 1989228477-A/1
PD 12-SEP-1989
PF 09-MAR-1988 JP 1988053774
PI AKINO TOSHIO, HORIKOSHI KOKI
PC C12N15/00,C12N1/20//C12N9/24,(C12N15/00,C12R1/07),(C12N1/20,
PC C12R1/19);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain-AM-001;
FH key Location/Qualifiers
FT CDS 1..1461
FT /product="beta-mannanase of Alkalophilic FT
FT /location/Qualifiers
FT /organism="unidentified"
FT /db_xref="taxon:32644"
BASE COUNT 475 a 257 c 340 g 389 t
ORIGIN
Query Match 7.4%; Score 109.4; DB 5; Length 1461;
Best Local Similarity 55.0%; Pred. NO.2.3e-15;
Matches 259; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

OY 1000 tctgtaggaagactcccgacacactcttatgatttgaaggtatgacaaagattg 1059
DB 991 TATGTCATCAATCAATCCAGCCACTGTTCTATGATTTTGAACAATACGCTATCGTGG 1050
OY 1060 actggaagtagctt---gagcgaggtccctgggctgtgacaaaggttgcttctaaagga 1116
DB 1051 TCCGGGTGGAATTCACGACGAGACCATGACTTGAATGATGCTGCGCAATGCT 1110
OY 1117 agtcattctttaaagcgatattcaattgtgtcaattcacacattacttactgtt 1176
DB 1111 ACTCATGCTGTAAGACGATGCTGTTCTGGCAATTAATAGTACCAT---TTGCAAAAA 1167
OY 1177 attcaaatagctctttaaagcagaatagtagtaacaaagctactgtttaaactgcaat 1236

```

```

DB 1168 ACAGTAATCGAATCTTACTTCATTCATAAAACCTAGAAATTAAGTACGACATCTTCG 1227
OY 1237 tgggaagtggttgtaaatgaaatgactgcgctcttctatgaaacggacatggtat 1296
DB 1228 TGGGAAATGTAGGAAGTGGCATGACAGCAAGATTTTCGTCAAAACGAGGAGTCTTGG 1287
OY 1297 acatgtactctggaaagcttctgtccgattaacggttctcctggacaacagctactcta 1356
DB 1288 AGATGAAATGCAAGTAATTTGTCAGTTTGCAAGCAACGAACACGACATATCTAAT 1347
OY 1357 gattatcaaatgtccaaaactcttcacaagtaagaaatgtaggttcattcaatca 1416
DB 1348 GATTGACGAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1407
OY 1417 gcgagtagatagtagtgcacaaatcatttatatgataatgattgta 1467
DB 1408 CCAGCAATAGCAACGAGGAGCGGATTTACTTATGATCAATGTGACGCTA 1458

RESULT 8
BACMANN 1939 bp DNA BCT 26-APR-1993
LOCUS BACMANN Bacillus sp. beta-mannanase gene, complete cds.
DEFINITION M31797
ACCESSION M31797
VERSION M31797.1 GI:143166
KEYWORDS 1,4-beta-D-mannan mannanohydrolase; mannan
endo-1,4-beta-mannosidase; mannanase.
SOURCE Bacillus sp. (strain AM-001) DNA, clone pMAH5.
ORGANISM Bacillus sp.
REFERENCE 1 (bases 1 to 1939)
AUTHORS Akino,T. and Horikoshi,K.
TITLE Two bacillus beta-mannanases having different COOH termini are
JOURNAL Appl. Environ. Microbiol. 55, 3178-3183 (1989)
MEDLINE 90146329
FEATURES
source location/Qualifiers
1..1939
/organism="Bacillus sp."
/db_xref="taxon:1409"
220..297
/note="beta-mannanase signal peptide"
220..1761
/note="beta-mannanase precursor (EC 3.2.1.78)"
/codon_start=1
/transl_table=11
/protein_id="AA022586.1"
/db_xref="GI:143167"
/translation="MKYKKVAVYMAFIHFVSLPTISMSSEANGAALSNPNANOMTX
NVYSWLANLPKRSKRVSGHFGYISDTIAMTKQCARBELTGKRPGLSCDYKNWTP
LYVADQISYGNDELINFMNQGIVTISHPMPGFRSGENVTITLPSOFQVLMK
TTEGRKMDLDMKADGLDELONNGVTLRPLHEMGEFEMWGAEGYNPDY
YISARMDYOFYERKLNLDIIVYSPDYRDHVTYSYPGANVYDIVALDST
LTQYNRIALDKRFAPAEIIGPPESMAGSDYSNYIOAKOKPRIYVLE
HNNGAWDLENDVYVNRGELIDYQSPATVIVDFENITLSMGSCEPTDGERISN
SANGTQSKADVYVGNNSYHOKTVKNKLSFRULEKVKVSHSGANGSGMTSVYFY
TGSAMRNWAGEFCOFAGKRTTALSIDLTKVSNLHDVREIGVEIKAPANSNGKRIYV
HVTVR"
298..1758
/note="beta-mannanase"
1826..1840
/note="pot. transcription termination signal; putative"
1861..1889
/note="pot. transcription termination signal; putative"
1905..1916
/note="pot. transcription termination signal; putative"
BASE COUNT 637 a 343 c 420 g 539 t
ORIGIN 1 bp upstream of XbaI site.
Query Match 7.4%; Score 109.4; DB 2; Length 1939;

```

Best Local Similarity 55.0%; Pred. No. 2.2e-15;
Matches 259; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

QY 1000 tctgtaggaaggaactctccgacacactcttattgtaagtagtatacgaagatg 1059
DB 1288 TATGTCATCAATCAATCAGCAGCTCTCTATGATTTTGAATAACATAGCATCTG 1347
QY 1060 actggaagtagctt---gaagcgagtgctctggctgtgacagagtgcttctaaga 1116
DB 1348 TCCGGGTGTAATTTACGAGGAGGACCATGACCTTCGATGATGCTGGCAAAATG 1407
QY 1117 agcatctcttaaaagagagatattcattgctgtaacatacagacttaactatgct 1176
DB 1408 ACTCAATCGTTGAAGAGATGCTGTTGGCAATATAGTCATCAAT---TTCGCAAAA 1464
QY 1177 attcaaaatacgcctcttaccagcagaatagtaggatacagaactactgttaacatgcaat 1236
DB 1465 ACAGTGAATCGAATCTTATTCTATTCAAAAACCTAGAAATTAAGTACCATTTCTTCG 1524
QY 1237 tggggaagtgctgttaagtaagtaagtgctgctcttcttctgaaacagagacatggtat 1296
DB 1525 TGGGGAATGTAGGAAGTGGCATGACAGCAAGAGTTTTCGTAACACGAGAGCTTGG 1584
QY 1297 actgtgactctggaagctgttgccgataaagtgctcattcgtgaacagactatctcta 1356
DB 1585 AGATGGAATCGAGTGAATTTGTCACTTTGCAAGGCAACCAACCGACATCTTATT 1644
QY 1357 gattatcaaatgtccaaactcttctcaagtaaggaattgagtgcttcaatca 1416
DB 1645 GATTTCGCAAGTAAGTAACTATCTGCATGATGTCGAGAGATAGTGTAGATATAAGCA 1704
QY 1417 gcgagtagatagtagtgacaaacatcgattatattgataatgtagtattga 1467
DB 1705 CCAGCAATATGCAACGAGGAGAGCGGATTTACTTATGATCATGACGACCGTA 1755

RESULT 9

SC2H4 25970 bp DNA BCT 19-OCT-1999
LOCUS Streptomyces coelicolor cosmid 2H4.
AL031514
VERSION AL031514.1 GI:3559956
KEYWORDS aminotransferase; beta-mannosidase; permease; secreted
beta-mannosidase; sugar transport system; sugar-binding
lipoprotein.

SOURCE

ORGANISM Streptomyces coelicolor A3(2).
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE

1 (bases 1 to 25970)
Kinschbach, M., Klesner, H.M., Denapate, D., Elchner, A., Cullum, J.,
Riedel, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)

FEATURES

source

CDS are numbered using the following system eg SC2H7.01c, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complement),
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in *S. coelicolor*
using a specially developed Hidden Markov Model (Krogh et al., 1994).
Nucleic Acids Research, 22(42):4768-4778(1994)) and the EMBL
program of Bibo et al., Gene 30:157-66(1984) as implemented at
<http://www.nh.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, ggg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid 2H4 lies between ZG5 and AH10 on the AseI-B genomic
restriction fragment.

location/Qualifiers

1. 25970

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid 2H4"

1. 1822

/gene="SC2H4.01"

<1. 1822

/gene="SC2H4.01"

/note="SC2H4.01, probable ATP/GTP binding protein, partial
CDS, len: >606 aa; contains PS00017 ATP/GTP-binding site
motif A (P-loop)"

/transl_table=11

/product="putative ATP/GTP binding protein"

/protein_id="CA20595.1"

/db_xref="GI:3559957"

/db_xref="SPTREMBL:O86584"

/translation="IDSTAPKAPVITSLSPYOCNTADCEKGGPGTAGFTPEPNA
DDIDNGYRWRLTTSKAKAYTGKSVYSDPSISGTOVLSVEAKDVRWGPAPAE
FEFKYAPAGATGTHFDALPDSGTTAKDTPABETRDHATLYTGAGKSTLARGD
AONSLWLDNGSDOTQAGYAATSAFVNTKDSFTVSAMAYLTDTSQTRVMAAPGTG
SAFTLYYSASYKRWENRAGVDRKPYRLSLGATAPLPLVWTHLAAPFKDNTN
KANDTQLFVNGRPGPEVLDGVSAYQPMVSGSLQIGSLVKGAGMEHFRGRIDE
VWVWQVDFPDELTQOALLEGVAPANEVLAQMDAASGTTVKEKLSVPAPSLTSA
ACAVYDENNALVILDSAGYASGATGVNDETFSPFYSVAQVDSKAKSPGVETLV
ACQASAGSSNALNAKVPADSVYQKRTTRIVAGADKVTQSAEAPGCDIATDTVQ
VTGVSDAQPWEMETPDASAKETRTGRILVGVETDQSEPTTPOYGAGALAV
GSGSGGTGGHPLGALIELRWVTGMSADQVRSOVLGCV"

188. 211

/gene="SC2H4.01"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

1949. 1953

/note="possible RBS upstream of SC2H4.02"

1959. 8510

/gene="SC2H4.02"

1959. 8510

/gene="SC2H4.02"

/note="SC2H4.02, unknown, len: 2183; very limited
similarity to proteins of the RBS family e.g. WAPA_BACSU
wall-associated protein precursor (2334 aa), fasta scores:
opt: 141 z-score: 296.8 E(-): 2.7e-09, 23.4% identity in
1237 aa overlap, and RBSA_ECOLI_RBSA protein precursor
(1377 aa), fasta scores: opt: 135 z-score: 222.4 E(-):
3.8e-05, 23.0% identity in 1016 aa overlap. Contains

RBS

gene

CDS

misc_feature

| Db | 25173 | ACGGCCACGATCCCTGGGCGCTCGGCTACCTCGGCTGTGCTCTCGAGGCGGCAACGGCGCGGCG | 25232 |
|----|-------|------------------------------------------------------------------|-------|
| Qy | 871 | tgggagatattttagaccttcgaatgatgttggctgtgaaataaccttcagcttggggaat | 930 |
| Db | 25233 | GTGGAATACCTGCAACATGTGTCAACGGCTTCGACCCCACTCGCTGACGAGCTGGGGCAAC | 25292 |
| Qy | 931 | acatagttaatggtcatatgttttaagaagaacttcgagattaagcacggtttaca | 990 |
| Db | 25293 | CCGATCTTTACGGCGACGCAACGGCATCGCCGCGAGCTCCAGAGACGGCCACCTGTACGGC | 25352 |
| Qy | 991 | ggtgagag | 998 |
| Db | 25353 | GGTGTGG | 25360 |

[illegible]

JOURNAL MEDLINE 97000351
 REFERENCE 2 (bases 1 to 14922)
 AUTHORS Seeger, K. J. and Harris, D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 14922)
 AUTHORS Parkhill, J., Barrell, B. G. and Rajandream, M. A.
 JOURNAL Direct Submission
 Submitted (04-OCT-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
 Streptomyces coelicolor sequencing at the Sanger Centre is funded by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
 CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(12):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
 Cosmid F73 overlaps with cosmid F11 on the AseI-F genomic restriction fragment.

FEATURES
 source 1. 14922
 location/Qualifiers
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid F73"
 1. 122
 /note="nominal overlap with Stf11"
 complement(1..983)
 /gene="manA"
 complement(<1..983)
 /gene="manA"
 /note="SCF73.01c, manA, probable secreted beta-mannosidase, len: 327 aa; similar to many e.g. MANA, STRLI P51529 mannan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363 aa), fasta scores; opt: 1189 z-score: 1332.7 E(): 0, 61.1% identity in 288 aa overlap. Contains N-terminal signal sequence, and Pfam match to entry PF00150 cellulase, Cellulase (glycosyl hydrolase family 5). Also similar to SC2H4.16 (62.2% identity in 288 aa overlap)."
 /codon_start=1
 /transl_table=11
 /product="putative secreted beta-mannosidase"
 /protein_id="CAB57406.1"
 /db_xref="GI:6013077"
 /translation="MRPARRDTPTPLPARLTLGLALGLVVGALCPGALASPAA

RBS
 gene complement(1129..2724)
 /gene="SCF73.02c"
 complement(1129..2724)
 /note="SCF73.02c, possible membrane-bound oxidoreductase, len: 531 aa; some similarity to eukaryotic molybdenum-containing oxidoreductases e.g. SUOX from *Sulfolobus solfataricus* (488 aa), fasta scores; opt: 2-score: 281.9 E(): 2.5e-08, 30.4% identity in 319 aa overlap, and NIAL_ORYSA nitrate reductase 1 (916 aa), fasta scores; opt: 247 z-score: 262.1 E(): 3.1e-07, 28.9% identity in 336 aa overlap. Contains several membrane spanning hydrophobic domains near N-terminus, and binds molybdopterin binding domain"
 /codon_start=1
 /transl_table=11
 /product="putative membrane-bound oxidoreductase"
 /protein_id="CAB57407.1"
 /db_xref="GI:6013078"
 /translation="MSDDKTPRNPSTARTRGALSGLLAGALAVELAAVR PRSGPVAVGGAAIDRTPTAVKDAIRFGINDVLTGLAVTLFALGLGFAVR HRRAGAGVLFVGVGAALAGALSGSTGTPSPVGAIGAVLALGLADLPAPK EAPERGGMRRRPRVATAAAGAGTGVGLRAGSAGREAVASSENTLEPPS RARVPRAQARVGVSTPTPTGDFVRDVALVIVRDLALMECRPRPSAGRNL TFDDLRLKRLERITTLTGVSNVGVGPRVGNRIGRDLALMECRPRPSAGRNL QLVARSVDGMTGSPVEDVMDGRUAMLAAGNGLPLRDFGVRAVVPGLVGVSAK RMIEDIELTFPDSDAYVWRKMARREAPVKQSIDTPRPARAGVAVVAGVAAQ HRGIDKVEVRDQDPEAEVLAADSBDTMRQMSVAMRATKGTTLVRADPRGEVQ TDKRRTRYPCGASRHSVYVVD"
 complement(12777..4300)
 /gene="SCF73.03c"
 complement(12777..4300)
 /note="SCF73.03c, unknown, len: 507 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF73.03c"
 /protein_id="CAB57408.1"
 /db_xref="GI:6013079"
 /translation="MTAFILVSGMFTGHIWMDTAAARLTARGAEVHTVATGLDAPRA AARGVDLFEHIALDVAVDSVGAAGRRIVLVGHDIHPVAGAAARRARVRIVH LDSGLPRDGVFALAAVDPQSLRELRAGAAGDVLPPRAHEPRKSTAGVPAALD RLIALAQPGLSTLQPLRLTGAVDPVPTGVLTGTGNTSELVQMLRLGLDPLRLP TDPVRSFEELPTGHPMLSLPAELTDLVLRAAAGRLEPVDVDTAGAGLRLPMDV PDVPRERGNLDLTPDAEPRPAVVVHGGVPPADARPTPRDMPGLTGLRVCVAGG AVGALDLRLHLDGDFERAAADVAAVEAVADRDVDAWDFEFGCGCLADMDL APAMIRGICATVTPVLAIPRMVGSSESPRVRVAVNAGALPVVLTTRGMRPIAT VEEFLAAKCGAGVEVVDVPHGHGFTIDLTDESRVAVHALRTVLDARFGHAR"
 complement(4311..4315)
 /gene="SCF73.04"
 4382..4387
 /gene="SCF73.04"
 4392..5414
 /note="SCF73.04, probable transcriptional regulator, len: 340 aa; similar in N-terminus to members of the MerR family e.g. NOXA_BRN8 nodulation protein NOXA (237 aa), fasta scores; opt: 211 z-score: 233.9 E(): 1.2e-05, 48.5% identity in 66 aa overlap. Also similar to SC017.06c (34.8% identity in 322 aa overlap). Contains Pfam match to entry PF00376 merR, Bacterial regulatory proteins, merR family"
 /codon_start=1
 /transl_table=11
 /product="putative MerR-family transcriptional regulator"
 /protein_id="CAB57409.1"

```

/db_xref="GI:6013080"
/translation="MFTSNDLGGICGLAAGAVTYKTVFVSDRLDLPASRSAGH
RRYGPEALDRRLRLRSJLALGLVPEVRLILDEHNAHTGGGCGALBAVGRLEEV
GSELRALMREARALRYOEAHPGQADRLRLIGAVSPSTALVFRWKMFAARPA
RSAGAFIDVAVPPDEPPAPPAVQVLAFLNALTLAPRPTVPOPPAHAAARAPA
LLYGLAEVLAAGVHMRGRGEPGALAFVASYSAVGAVDPDQFRRRLAQLAA
DPRIDGVELVAEYTTAPGARPEPTPSADHMLAALTAGAVDPDQFRRRLAQLAA
SPGLGV"
gene
complement(5357..5623)
/feature="SCF73.05c"
complement(5357..5623)
/feature="SCF73.05c"
/feature="SCF73.05c, unknown, len: 88 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.05c"
/protein_id="CAB57410.1"
/db_xref="GI:6013081"
/translation="MDEVEFMGRVYAGADHNPGRAGRYAHLVGGFLDLIDVTD
LTEORARGVTLATGIRYGPCHSAYTPRPGDTRFRDVRDVP"
complement(5696..6697)
/feature="SCF73.06c"
complement(5696..6697)
/feature="SCF73.06c"
/feature="SCF73.06c"
/feature="SCF73.06c, unknown, len: 333 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.06c"
/protein_id="CAB57411.1"
/db_xref="GI:6013082"
/translation="MLNDPVTVDKGRYDYTPMEEPARTTEALGNTAEPLAAGLR
PTGRMVRILPMPVILRLCVAAGDAVWFASPEAGLFEALASTARVPGVLEPLA
VDARGMRLPDPVLPFDVLAEPDPAWELRLQYAMHALTSRSETERGAPV
AVAGAGAGVLDLDTAPLSARDRLRLDMLCVELALGVPSLDIADHDG
QVPRPARGPTFRPDMDVAVTGPFFSLAVPARAARVPOVLPRLDVALEPMTGEG
RTAQLRAVSLMRLSALNRAAYGRILPFGASGAGATTAAGARCLLELDEPPL"
complement(6778..9000)
/feature="cpeb"

```

Query Match 3.6%; Score 52.4; DB 2; Length 14922;
 Best Local Similarity 46.9%; Pred. No. 0.045;
 Matches 275; Conservative 0; Mismatches 296; Indels 15; Gaps 3;

```

OY 101 cggattatgtaagcgggtaccctctatagccatgtaaacatccattgtaata 160
DB 814 CCGGCTGCACATCGGTACGCGCCGCTGCGAGCGGACGACACTCTGCTATGC 755
OY 161 gaggatcaaccatgagcagcgtgtaaaagaccagcactactgcaattgaaggaa 220
DB 754 GCGGGGTCAACCGCCACCTGATCCCGGGAGAACCGAGCTGGCGGACGTCA 695
OY 221 ttgcaaatccggtgtaatacgcgtggtgtgttctctgtaggggaaatggagaa 280
DB 694 AGGC--GCTGGCGCGAAGACGCTCGGGGTGCTCTCCGACGCGACCGCTGAGCG 638
OY 281 aagatgaacatcatagcaagaacactatctcttagcgggaagataatcatctgtg 340
DB 637 AGAACCGCCCGGCGAGCGCGCGCGCTATGAGCAGTGCAGGCCAACCGGCTCATCT 578
OY 341 cgtgttgtaagtctgtagcagcgtatgattcattgc-----ttgcctca 391
DB 577 GCGTACTGAGGTGACGACGACACGCGGTACGCCGAGACCGCGCGGCGACGCTCG 518
OY 392 atcgtcgtgtgattttagatgtaaatgaaatgcttaattgaaaggaataacg 451
DB 517 ACCACCGCGCGGCTACTGATGCGCTCAAGAGCTCTGCGCGCGCGACAGGAGACTACG 458
OY 452 tctatttaattgtagaatggttggctcgtgggaaggaggtgcttggctgacg 511
DB 457 TCATCGTCAACATCGGAGAGAGCCCTGGGCAACCGCCCGCGCTGGACGAGAC 398
OY 512 ggtataacaagaatcccgatgctgtaacgctgtaaacatccattgattggtg 571

```

```

DB 397 CGACCGTCCGCGCGCTCAAGACGTGCGCGCGCGCGGCTCCACACAGCATCATGTCG 338
OY 572 atgtcgggggtggggacaattccacaatg--attcagattatggaagaaattt 628
DB 337 ACCGCCCCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 278
OY 629 ttatgtcgcctcaacgaataatgatttgcattatgta 674
DB 277 ACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 232

RESULT 13
LOCUS 166494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
location/Qualifiers
source 1..7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

```

Query Match 3.3%; Score 48.6; DB 5; Length 7218;
 Best Local Similarity 3.8%; Pred. No. 0.35;
 Matches 15; Conservative 216; Mismatches 160; Indels 0; Gaps 0;

```

OY 747 agtcattgtaattggaacccgtcatcaacaatgagtgatgtaagcaacattat 806
DB 1455 AGAGATAGAGAAATTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1396
OY 807 ggcattctgaaacaagaagatggtggtgctgctgcatgaaaggaagcgcc 866
DB 1395 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1336
OY 867 agaatggagatttagacccttcgaatgattggcgtggaataacttaagcttggg 926
DB 1335 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1276
OY 927 aatacaatgtagatgctccatgcttgaagaaactcgaataatgaacacgctt 986
DB 1275 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1216
OY 987 tacagtgagagatcgtatgaggaactctccgaacactcttagatttgaagtag 1046
DB 1215 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1156
OY 1047 tatgcaagatgtagcgtggaatgactctgagcggaggtccttggcgtgtagaagtag 1106
DB 1155 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1096
OY 1107 ttctaaaggagatcattctttaaagcggat 1137
DB 1095 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1065

RESULT 14
LOCUS AC024591 111071 bp DNA HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-511G21, WORKING DRAFT
SEQUENCE, 13 ordered pieces.
ACCESSION AC024591
VERSION AC024591.3 GI:9256450
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 11071)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 16
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 11071)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
COMMENT Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7705016.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 602754
Center Clone name: RPCT-11_511621

Summary Statistics
Consensus quality: 104021 bases at least Q40
Consensus quality: 108348 bases at least Q30
Consensus quality: 109319 bases at least Q20
Estimated insert size: 118910; agarose-1p estimation
Estimated insert size: 110521; sum-of-contrigs estimation
Quality coverage: 8.44 in Q20 bases; agarose-1p estimation
Quality coverage: 9.08 in Q20 bases; sum-of-contrigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contrigs. Gaps between the contrigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
3999: contrig of 3999 bp in length
4000
4099: gap of unknown length
4100
6224: contrig of 2125 bp in length
6225
6324: gap of unknown length
6325
14602: contrig of 8278 bp in length
14603
14702: gap of unknown length
14703
54245: contrig of 39543 bp in length
54246
54345: gap of unknown length
54346
58569: contrig of 4224 bp in length
58570
58669: gap of unknown length
58670
59480: contrig of 811 bp in length
59481
59580: gap of unknown length
59581
60683: contrig of 1103 bp in length
60684
60783: gap of unknown length
60784
62950: contrig of 2167 bp in length
62951
63050: gap of unknown length
63051
64092: contrig of 1042 bp in length
64093
64192: gap of unknown length
64193
85067: contrig of 20875 bp in length
85068
85167: gap of unknown length
85168
91149: contrig of 5982 bp in length
91150
91249: gap of unknown length
91250
108236: contrig of 16987 bp in length
108237
108336: gap of unknown length
108337
11071: contrig of 2735 bp in length.
Location/Qualifiers
1. 111071
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-511621"
/clone_lib="RPCT human BAC library 11"
BASE COUNT 27517 a 27152 c 28256 g 26944 t 1202 others

ORIGIN
Query Match 3.1%; Score 45.8; DB 53; Length 111071;
Best Local Similarity 34.7%; Pred. No. 1.5;
Matches 110; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
QY 322 gaagataacatttggtcttcttgaagttcaatgaatgcacggttatgattccatt 381
DB 59292 GATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59351
QY 382 gcttcgctcaatcgtctgctgctgctgctgctgctgctgctgctgctgctgct 441
DB 59352 GATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59411
QY 442 gaagataacgctcattataataatgcgaatgaatggttgcgtgctgctgctgctgct 501
DB 59412 GATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59471
QY 502 tgggtcgaacggtatataaacaagcaatcccgatgctgcaagcggtctaaccctacc 561
DB 59472 GATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59531
QY 562 ttgatgtagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 621
DB 59532 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 59591
QY 622 gaagtttcaatgctga 638
DB 59592 GATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59608
RESULT 15
LOCUS CELK03B6 35049 bp DNA INV 28-OCT-1998
DEFINITION Caenorhabditis elegans cosmid K03B6.
ACCESSION U55375
VERSION U55375.1 GI:3805671
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
REFERENCE
AUTHORS 1 (bases 1 to 35049)
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copesey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirshen, J., Lalister, N.,
Lattelle, P., Lightning, J., Lloyd, C., McKerray, A., Mullimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Riken, L., Koop, A.,
Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, K., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, R., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
REFERENCE MEDLINE 94150718
AUTHORS 2 (bases 1 to 35049)
Lattelle, P. and Gattung, S.
TITLE The sequence of C. elegans cosmid K03B6
JOURNAL Unpublished (1998)
REFERENCE 3 (bases 1 to 35049)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (17-Apr-1996) Robert Waterston
REFERENCE 4 (bases 1 to 35049)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-Apr-1997)
REFERENCE 5 (bases 1 to 35049)
AUTHORS Waterston, R.

TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On Oct 28, 1998 this sequence version replaced g1:1280123.
Submitted by: Genome Sequencing Center,
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwenemate@wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is C04D1, 2700 bp overlap; 3' clone is P49E7, 200 bp overlap. Actual start of this clone is at base position 701 of CELK03E6; actual end is at 35049 of CELK03E6

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

1. 35049

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="K03E6"

/chromosome="X"

/complement(3691..7748)

/gene="unc-1"

complement(join(3691..3766,3931..4145,4822..5095,5154..5278,6432..6545,7296..7326,7735..7748))

/gene="unc-1"

/note="K03E6.5"

/codon_start=1

/product="Erythrocyte band 7 integral membrane protein"

/protein_id="AAC69046.1"

/db_xref="GI:1280125"

/translation="MEYSHQIPKSCATIDVPPDYETIGTFGLALMSLIIIVF

PSMCVCLVKEVERVYRIGLVFGARPGMIFIPIDYRKIDLVSVIAP

PEITSKDSVTVSDAVVYFRTSDPIASVNNDDAIVSTKILADITLNNALMTLE

MTERAIAIOLCEITIDECHEWGVKVEVKDRIEQLQTRAAAEARARARAK

VVAASEQKASRAIKRAADYIQANVVALQRLQLALNSIAEHNSTIYFVPVMEFGA

FMKDKQ"

complement(12649..14893)

/gene="K03E6.4"

complement(join(12649..12750,12801..12928,13397..13527,13570..13666,14217..14391,14837..14893))

/gene="K03E6.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69043.1"

/db_xref="GI:1280126"

/translation="MERADSVILTIHGNVKTTFEPRESVYKLTSTVATGPKIINGE

TVLSISFSGADNOIFIDPEFGKGSPPMPLVINCFSKHSKDCVGTIGML

PIPEPEKHYCHVHCFLPEPSIVQOILALRVNAFVDPMLPANAADRYACRTSTK

GYVDLEINVTIKCSDRPKRTSTSTELKSSNSHATNNVSTTIADLIGDAVANEETFN

AOEDNEESKR"

gene

20796..24499

/gene="lim-6"

join(20796..20881,21727..21835,22141..22262,23465..23577,23885..24151,24209..24499)

/gene="lim-6"

/note="K03E6.1"

/codon_start=1

/product="similar to other homeobox domains; contains two LIM domains"

/protein_id="AAC69042.1"

/db_xref="GI:1280127"

/translation="MSLLISATYSSTEDRLKSGCCGLIKOREFLIRTESKQMA

HLIADIFILITVLDKSNMDESYHSCRLSCCLSLSSFKGSRGNIGCEH

DRHMLYKRCRCMTLLPTDVRHVMHYHAOCFSCSCORFNDDEYHVDGEV

FCNDKQSLCNCQTSSEPHYTLFTTFVAVSVSPDIMEVEVSEIRKTPKRPITLN

ADQROKTAPESSKPSKRYKVEQALNETGVSRYVYVAFONRAKIKLKKDSDG

DFKFGSGSGSTEDIRSDDEESVLSKLRIQIDIGELMLYKSFQVKRFVP

"

complement(25406..26485)

/gene="K03E6.6"

complement(join(25406..25503,25556..25679,25726..25773,26375..26485))

/gene="K03E6.6"

/note="Similar to Profilin; K03E6.6"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69045.1"

/db_xref="GI:1280128"

/translation="MSMSDIINNLLIGSGVSKAAILFGDAVMAKSDNFNISVEAV

AAGAFSLDALGTGIRLEQKEVFLNMDNDILIKQSGSGFIKTIQAVIISYE

KGLQPEMCSSTGALADYRSIKY"

complement(28061..33962)

/gene="K03E6.7"

complement(join(28061..28224,28960..29077,30844..30997,31048..31205,31305..31629,32214..32347,33211..33354,33555..33654,33708..33861,33914..33962))

/gene="K03E6.7"

/note="similarity to C. elegans protein C01C10.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69046.1"

/db_xref="GI:1280129"

/translation="MDEATSEVTEVFSDPTTRPFEDGHPELETRHLMTHLELE

KLINSTDVINKMEYOLADRADPEROVONSEKLEKSOYSQIAKAPFYELIKE

BSLRESEKAEERBERATSTIGIKOOVSLTOSLSOTSLPECLEVLNHHIOVRE

VEERTAESLHASNAHMLAKIKIAMENDNATYKRSIYVEKLEPTEKILEAK

AIILCELEAVQKNDYTSUNLERISERIERHRSRGSSEAVSSPOEDOKSPKES

ESIPGNPPVAPATPPYEDKYIIDKDDSVLNMNITDEDEBEKRSNRSISGVIL

LAOOLIGNGNSTEKHNTIPRHGEADISYHTRVAGSDSDNSSEVSLASFNIGDL

TVSKMLMSHELKECAIKITEAIVAVAKFONRQLPKIQVADKRLRPTPPCKQ

IKNSLTSLDPLFANOSIVFEISMAVGSGYPPFLSLCEFRVSGTCR"

BASE COUNT 11644 a 5816 c 6271 g 11318 t

ORIGIN

Query Match 3.1%; Score 45.2; DB 33; Length 35049;

Best Local Similarity 51.6%; Pred. No. 2.1;

Matches 128; Conservative 0; Mismatches 118; Indels 2; Gaps 1;

196 caagcaactacgcaattgaaggattgcaaacacggcggtcaatacgggtccggattg9 255

DB 33790 CATGAATCATATTCATCAATGACGAGTTCCAGTGAGGAGGTCATCTTTCGAATTCG 33810

QY 256 ttatctgatgggggacacatgacaaagaatgacatcagtaagaacattatctt 315

DB 33850 TTGTTGGGGGCTCTGGAATTCACAAATTCGCAATTTAGAAATTAGAAATAGATGCT 33900

QY 316 ttaacgaagaataacattggttcgtctcctgaagttcatgatcaccggtatga- 374

DB 33910 TTACTGAAACAAATGTTTCGTCGACCTCCGAAAGAGTTTCAGTTTCGTCATTTTGA 33969

QY 375 -ttcatggttcgtcgaatcgtcgttgaattgatgataagaatgaagtccttaa 423

Db 33970 CTTTATAGGTCATGCAATTTTGTGGTGTCTTTGTTCTGGAAGTAAACATTTTAT 34029

OY 434 ttggaag 441

|||||

Db 34030 GCGGAATG 34037

Search completed: December 20, 2000, 03:12:31
Job time: 12796 sec

This Page Blank (uspio)


```

QY 902 ctggaataaacttacagcttggggaatacaatagtgatggtccatagtgttaag 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 902 ctgttaacactttaactgacttggggaatagaaattgtcccgggcgagtgcttaag 961
QY 962 aaacttcagagatgaagcagcttttaacaggt-----ggagatcgtatggaggaact 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 962 aaacttcagagatgaagcagcttttaacaggttttaacaggttttaacaggttttaac 1021
QY 1016 ctccgaacactctttatgatttgaagtagtataaggaatgagtgagtgagtgagtgag 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1022 ctgtactactactgtatgacttgaaggaagcagacaaaggtgcatgaggaagcaagctga 1081
QY 1076 gcggaagctcttggctgtgacagagatggtctcttaaggaatcattctttaaagcg 1135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1082 ccggttgccttggctgtacagagatgggtgtcttaagtaactctctttaaagcg 1141
QY 1136 atattcaattgtctgaacttcaacattactactatgatttcaaaaatagctttac 1195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1142 atgtaatttaacctcaaatcttcaacatgaaactgataagcaaaagcgttaactac 1201
QY 1196 agcaagatagtagaatacaagctactgtttaacatgcaaatgggggaagtgtgtaatg 1255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1202 acggaatacttcagacagcaagcagcttcgcatgccaattggggaatcccggaatg 1261
QY 1256 gaatgactgcgctcttattgtgaaacagagacatgttatacatgtactctggaagct 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1262 gcattgaatgcaagactttaaactgaaacgggtctgtattatagcatgcatgagtgctct 1321
QY 1316 ttgtgcgataagcttgc---atctggaaacagcgtatctcaggttttcaaatgtcc 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1322 ttcaacgtatacaatagctccaactcaggaacagcgtatcttatttaacaacacag 1381
QY 1373 aaaactcttctcaagaaggaatgtgagtgatgtccatcagagagtgatagtg 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1382 aaatagtcatactgttagggaatagggcgtgaatttcaaggcgaataatagcagtg 1441
QY 1433 gacaacatcgatttatattgataatgtgattga 1467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1442 gtcaactgtctatactatgatacgttacttta 1476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
245337
ID 245337 standard; DNA; 1482 BP.
XX
XX
AC 245337:
XX
XX
DT 27-MAR-2000 (first entry)
XX
XX
DE DNA encoding a Bacillus mannanase enzyme.
XX
XX
Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KM plant material degradation; recycled waste paper; paper making pulp;
KM guar; locust bean gum; mannan-containing food; coffee extract;
KM cleaning composition; machine washing; hard-surface cleaner;
KM dishwashing; oral; dental; contact lens; body-care composition;
KM fabric softener; oil well drilling; subterranean formation fracture; ss.
XX
XX
OS Bacillus agaradhaerens.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1482
FT FT /*tag= a
FT FT /product= "mannanase"
FT FT 1..93
FT FT /*tag= b
XX
XX
W0964619-A2.
XX

```

```

PD 16-DEC-1999.
XX
XX 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI MPI: 2000-105891/09.
XX
DR P-PSDB; Y54124.
XX
XX New mannanases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions
XX
XX Example 5; Page 212-213; 242pp; English.
XX
XX The present sequence encodes a Bacillus mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
XX mannanase hydrolyses galactomannans. Specifically, mannanases hydrolyse
XX 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
XX and galactoglucomannans. The mannanase protein, or preparations
XX containing it, are used to improve properties of cellulosic or
XX synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannanases
XX are also used to process coffee extracts (to inhibit gel formation);
XX in cleaning compositions (for machine washing of fabrics, as
XX hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX dental, contact lens or body-care compositions) where they remove
XX mannan-containing soils and prevent binding of some soils to
XX cellulosics; and in fabric softeners. They can also be used in oil
XX well drilling to fracture subterranean formations.
XX
XX Sequence 1482 BP; 446 A; 285 C; 352 G; 399 T; 0 other;
XX
XX
Query Match 44.0%; Score 646.6; DB 21; Length 1482;
Best Local Similarity 65.9%; Pred. No. 4,2e-171;
Matches 97/2; Conservative 0; Mismatches 494; Indels 9; Gaps 2;
QY 2 tgataatggtttaaaaaatttttcaatacatattactactattagtgactga 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 tgaataaaaaagtatacaagatttacttattattttggcacactataaagaagtg 61
QY 62 ttctgttcttcaaggaacttcaagcgaatgcaaatcccggaattttgtaagcggtga 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 gaataatgggattacaacagtcgccatcagcagaagtaacagcttttattgatagga 121
QY 122 ccactctatagatggcaatggaacccattgttaatggaggaattaaccttgagcag 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 atacgttataatgacgaagaatggcagcattgttctgaggaagtataacacgtgagcag 181
QY 182 catgtataaagcagcgaactatgcaattgaaaggattgcaaataccggttgacta 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 ctgtgataaagacaccgcttcaacagcgtattcccttgcatgacaggaagcgccaaca 241

```


PR 10-MAR-1999: 99US-0123623.
 PR 10-MAR-1999: 99US-0123641.
 PR 11-MAR-1999: 99US-0123642.
 XX
 XX (NOVO) NOVO-NORDISK AS.
 PI Kaupinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX WPI: 2000-105891/09.
 DR F-PDB: Y54125.
 XX
 XX New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PT
 XX
 PS Example 5; Page 214-215; 242pp; English;
 XX
 XX The present sequence encodes a *Bacillus mannanase* (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The present mannanase is a synthetic variant of the mannanase of
 CC Y54124, in which the C-terminus of the protein was changed due to
 CC design of a lower PCR primer used for amplification. The mannanase
 CC hydrolyses galactomannans. Specifically, mannanses hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanses
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 CC
 XX Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other;
 SO

Query Match 41.4%; Score 608.2; DB 21; Length 1407;
 Best Local Similarity 65.6%; Pred. No. 2,2e-160;
 Matches 921; Conservative 0; Mismatches 473; Indels 9; Gaps 2;

| | | | |
|----|------|--------------------------------------------------------------|------|
| QY | 422 | gaagtccttaattggaagaagatccgtcatttaataattgcgaatgagtttg | 481 |
| DB | 422 | aagatgccttaattggaagaagatccgtcatttaataattgcgaatgagtttg | 481 |
| QY | 482 | gttcgtggaaggagatgctgctgacgggtataaacaagaatcccgatgcgta | 541 |
| DB | 482 | ggaattggatgctgctgacgggtataaacaagaatcccgatgcgta | 541 |
| QY | 542 | acgcggtctaaacataccttgatgtatgctgcgggtggtggacaattcccaaat | 601 |
| DB | 542 | atgcggtctaaacataccttgatgtatgctgcgggtggtggacaattcccaaat | 601 |
| QY | 602 | cgattcagatgataagaagaattttaatgctgcacccccaagaataacatgttt | 661 |
| DB | 602 | ctatcagatgataagaagaattttaatgctgcacccccaagaataacatgttt | 661 |
| QY | 662 | cgattcagatgataagaagaattttaatgctgcacccccaagaataacatgttt | 721 |
| DB | 662 | cgattcagatgataagaagaattttaatgctgcacccccaagaataacatgttt | 721 |
| QY | 722 | gagtccttaacaaagactcgtcattgattgattggaaccggtacatacaatg | 781 |
| DB | 722 | gagtccttaacaaagactcgtcattgattgattggaaccggtacatacaatg | 781 |
| QY | 782 | gtgacgtcgatgaagcaacgattatgagctatttgaacaagaaggagttggtgtg | 841 |
| DB | 782 | gtgacgtcgatgaagcaacgattatgagctatttgaacaagaaggagttggtgtg | 841 |
| QY | 842 | cggttcagatgaagaaggagcccaagaatggagattgagcttgcgaatgattggg | 901 |
| DB | 842 | cggttcagatgaagaaggagcccaagaatggagattgagcttgcgaatgattggg | 901 |
| QY | 902 | ctggtctgtgaagaaggacagctacagatgagatattgagcttccagaagctggg | 961 |
| DB | 902 | ctggtctgtgaagaaggacagctacagatgagatattgagcttccagaagctggg | 961 |
| QY | 962 | aaacttcagatgaagcaacgatttacaagt-----gagagatctgagtggagaact | 1015 |
| DB | 962 | aaacttcagatgaagcaacgatttacaagt-----gagagatctgagtggagaact | 1015 |
| QY | 1016 | ctccgacaactcttattgatttgaagtagatgacagagatggaatgagttgagcttga | 1075 |
| DB | 1022 | ctggtcactccttctgattgatttgaagtagatgacagagatggaatgagttgagcttga | 1081 |
| QY | 1076 | gcggaggtccttgcgtgacagatggtctcttaaggaaagatcattcttaaaagcgg | 1135 |
| DB | 1082 | cggtggtccttgcgtgacagatggtctcttaaggaaagatcattcttaaaagcgg | 1135 |
| QY | 1136 | atattcaatttgcgtcaaatctcaacattacttactatgatttcaaaaatgctttac | 1195 |
| DB | 1142 | atgtaatttcaacccaatttcttcaacatgacttactatgatttcaaaaatgctttac | 1201 |
| QY | 1196 | agcagaatagtgatgatacagactactgttaacaatattgggaattggttgaatg | 1255 |
| DB | 1202 | acggaatcctcagctcaacgacacgcttgcgcacatggtgggaattcccgtaatg | 1261 |
| QY | 1256 | gaatgactgcgtcttcttattgtgaaacagagacatggttatacatgctcttgaaagt | 1325 |
| DB | 1262 | gcatgattggaagactttagtgaacagcgctctgattatacatgctcttgaaagt | 1332 |
| QY | 1316 | ctgtgcgattaaaggttctc-----atcggaaacagctatctgattatacaatgctc | 1372 |
| DB | 1322 | ctacagctatcaatagctcaactcaggaacagctatctgattatacaatgctc | 1381 |
| QY | 1373 | aaatcttctcaagcaaggaa | 1395 |
| DB | 1382 | aaatcttctcaagcaaggaa | 1404 |

RESULT 7
 Z45342
 ID Z45342 strand: DNA; 1107 BP.
 XX

| | | | |
|--------|-----|---------------------------------------------------------------------------|---------------|
| Db | 946 | ggtccgaatgggattcgctcaaacctctaaagaagcagctgtttt | 990 |
| RESULT | 9 | | |
| XX | ID | 245345 | |
| XX | | 245345 standard; DNA; 960 BP. | |
| XX | AC | 245345; | |
| XX | DT | 27-MAR-2000 | (first entry) |
| XX | DE | DNA encoding a <i>Bacillus</i> sp. mannanase enzyme. | |
| XX | | | |
| XX | | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | |
| XX | | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | |
| XX | | mannan; galactomannan; glucomannan; galactoglucomannan; cellulose fibre; | |
| XX | | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | |
| XX | | plant material degradation; recycled waste paper; paper making pulp; | |
| XX | | guar; locust bean gum; mannan-containing food; coffee extract; | |
| XX | | cleaning composition; machine washing; hard-surface cleaner; | |
| XX | | dishwashing; oral; dental; contact lens; body-care composition; | |
| XX | | fabric softener; oil well drilling; subterranean formation fracture; ss. | |
| XX | | | |
| XX | | <i>Bacillus</i> sp. | |
| XX | | | |
| XX | Key | Location/Qualifiers | |
| XX | FT | 1..960 | |
| XX | FT | /*tag- a | |
| XX | FT | /product- "mannanase" | |
| XX | FT | /note- "not termination codon given" | |
| XX | FT | 1..84 | |
| XX | FT | /*tag- b | |
| XX | | | |
| XX | PN | MO9964619-A2. | |
| XX | | | |
| XX | PD | 16-DEC-1999. | |
| XX | | | |
| XX | PF | 10-JUN-1999; 99WO-DK00314. | |
| XX | | | |
| XX | PR | 10-JUN-1998; 98US-0111256 | |
| XX | PR | 20-OCT-1998; 98DK-0001340. | |
| XX | PR | 20-OCT-1998; 98DK-0001341. | |
| XX | PR | 28-OCT-1998; 98US-0105970. | |
| XX | PR | 28-OCT-1998; 98US-0106054. | |
| XX | PR | 23-DEC-1998; 98DK-0001725. | |
| XX | PR | 05-MAR-1999; 99DK-0000306. | |
| XX | PR | 05-MAR-1999; 99DK-0000307. | |
| XX | PR | 05-MAR-1999; 99DK-0000308. | |
| XX | PR | 05-MAR-1999; 99DK-0000309. | |
| XX | PR | 09-MAR-1999; 99US-0123543. | |
| XX | PR | 10-MAR-1999; 99US-0123623. | |
| XX | PR | 10-MAR-1999; 99US-0123641. | |
| XX | PR | 11-MAR-1999; 99US-0123642. | |
| XX | | | |
| XX | PA | (NOVO) NOVO-NORDISK AS. | |
| XX | | | |
| XX | PI | Kaupinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME; | |
| XX | | WPI: 2000-105891/09. | |
| XX | DR | P-PSDB; Y54132. | |
| XX | | | |
| XX | PT | New mannanases for treatment of textiles, plant material and coffee | |
| XX | PT | extract, and in cleaning compositions - | |
| XX | | | |
| XX | PS | Disclosure; Page 227-228; 242pp; English. | |
| XX | | | |
| XX | CC | The present sequence encodes a mannanase enzyme (also known as | |
| XX | CC | mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). | |
| XX | CC | The mannanase hydrolyses galactomannans. Specifically, mannanases | |
| XX | CC | hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans, | |
| XX | CC | glucomannans, and galactoglucomannans. The mannanase protein, or | |
| XX | CC | preparations containing it, are used to improve properties of cellulosic | |

| | Query Match | 21.1% | Score 310.8 | DB 21 | Length 960 | |
|----|------------------------------------------------------------------------------|----------------|-------------------|----------|------------|--|
| | Best Local Similarity | 59.3% | Pred. No. 2.5e-77 | | | |
| | Matches 547 | Conservative 0 | Mismatches 372 | Indels 3 | Gaps 1 | |
| CC | or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based | | | | | |
| CC | sizes or printing pastes). They are also used to degrade or modify | | | | | |
| CC | plant materials (particularly recycled waste paper, paper making pulps, | | | | | |
| CC | or material containing guar or locust bean gums (chickeners), or to | | | | | |
| CC | reduce viscosity of mannan-containing foods or feeds). The mannaes | | | | | |
| CC | are also used to process coffee extracts (to inhibit gel formation); in | | | | | |
| CC | cleaning compositions (for machine washing of fabrics, as hard-surface | | | | | |
| CC | cleaners, for hand or machine dishwashing, also in oral, dental, contact | | | | | |
| CC | lens or body-care compositions) where they remove mannan-containing | | | | | |
| CC | soils and prevent binding of some soils to celluloses; and in fabric | | | | | |
| CC | softeners. They can also be used in oil well drilling to fracture | | | | | |
| CC | subterranean formations. | | | | | |
| XX | Sequence 960 BP: 257 A; 183 C; 280 G; 240 T; 0 other: | | | | | |
| SO | | | | | | |
| QY | 45 aactatagctagctctatcttctgttctgttccagaaactctacagctaatgcgaattccgg 104 | | | | | |
| DB | 33 actagtgtgtgtgtgtgttctgttctgttatacagtagtcggttttagatctgcgcaaaagcg 92 | | | | | |
| QY | 105 attatagtgaagcgtgtaccactctctacagatgcgaatggaaccattgtataagag 164 | | | | | |
| DB | 93 ctltacgctaaagtagacagagtgtgtgacaaaatgctgcattcccttaccgtatctgtg 152 | | | | | |
| QY | 165 gattacacatgagcgacgcatgtagtaaaagaccagacaactactgaatggaaggtatgc 224 | | | | | |
| DB | 153 cgtcaacatgtgacactctctgtgtttaaacagaatttagaggaaggaatccctgcatalgc 212 | | | | | |
| QY | 225 aaatccggtgcctaaacaggttcgcgagatctgtgtatctatgtatggtggagacaaatgcagaaaga 284 | | | | | |
| DB | 213 agaaacagagggcgacacacagtgtagatctgtcttaccatgtgacagcaatgtggaaaaga 272 | | | | | |
| QY | 285 tgaacatcacatagcagaagaacacttactctttagcggagaagaataacattgtgttcgt 344 | | | | | |
| DB | 273 tgaatgcctctgtgcttgcgcgtgtctgttctgcacagaaacatagtgttgacaacgct 332 | | | | | |
| QY | 345 tcttgaattcatagtgttaccggtttagatctcatcttcttcgcctcaatcgtgcctgtga 404 | | | | | |
| DB | 333 gctggaagctccacgtagtgcatacgaaggagataatcccgatgtattagataaaagcagctga 352 | | | | | |
| QY | 405 ttattgagattgaatgagaaagtctcttatttggaaaggaagaataccgtcatattcaat 464 | | | | | |
| DB | 393 ttactcgtgtcgaatgctgtatgttcttaaaaggagcaaaagccgggtcaatcattacaact 452 | | | | | |
| QY | 465 tgcgaatgaatgttgttctgttgcgtggaaagggtgactgttgcgtgcaggggtataaacaaac 524 | | | | | |
| DB | 453 tgcacatgaatgttagttagggcgtaggaagatgtagcgttttggcgaggggcatacgcaaacg 512 | | | | | |
| QY | 525 aatcccgagattgcgttaacgcgcgtgtctaaacatacctttagatgtgaatgtcggcggtg 584 | | | | | |
| DB | 513 gatcccgcgcttgcgcacatgtctggtccctgcgcacatcgttaagtctgttcgycgcaggtg 572 | | | | | |
| QY | 585 gggagcaattccacacatgcattcagatctatgtagaagagaagtctttaaagtcgcagccca 644 | | | | | |
| DB | 573 gggagagtaacctgcctctatccatagtagcggggagccgaagtatttgcctccgacatt 632 | | | | | |
| QY | 645 acgaaatacaatgttttcgattcatatgtatgaatatacaggttgttaatgcatcgcaagt 704 | | | | | |
| DB | 633 aaaaaacaatgtttccatccatcatatgtagcaatatgcaagtaggataaggcgacagct 652 | | | | | |
| QY | 705 tctgtactaatatitgacgagttcttaataacagacctgcattagtcattgtgtgaattgg 764 | | | | | |
| DB | 693 ttcttgaaaaacgtcgcggtgtgactgtctgtaaaaatcttctgtgtaaacoggtgaattgg 752 | | | | | |
| QY | 765 aacagcgtatcaaaatgttgaagtcgtgataaggaacaggttatagactcttccgaacaaag 824 | | | | | |
| DB | 753 ccatagtgatcatcagtgtggtcagatgtcogaagaagatgctgatttgcctataacgcagaaagc 812 | | | | | |
| QY | 825 aggaattgtgtgtgtgtgtgtcatgtgaaagggagaacgcccagaagtggaaattatga 884 | | | | | |

| Query Match | 12.0%; Score 175.8; DB 21; Length 564; |
|----------------------------------------------------------------|-----------------------------------------------------------------|
| Best Local Similarity | 58.2%; Pred. No. 1.1e-39; |
| Matches 309; Conservative 0; Mismatches 222; Indels 0; Gaps 0; | |
| QY 45 | actattagctagctctattcttctggttcgaagaaccttaagctaatgcaattccg 104 |
| DB 33 | actatgtagcggtgtgtgtttgtttgtatcacgtagcggttttagcatctgcacaaagcg 92 |
| QY 105 | atttattgaagcggtgaccacactctacgatgcacatggaacccatttgaatgagag 164 |
| DB 93 | ctttccgtaaaagtgatacagaagttgttgacaaaatgtagctcttaagttatgctg 152 |
| QY 165 | gattaaccaatggcagcagatgtataaagaccagcgcaactctgaattgaaggatg 224 |
| DB 153 | cgtcaacacatgacactcttctgttttaaaagatttagaggagcgcaactcctgtacag 212 |
| QY 225 | aaatccggtgttaataacgctgcggatgtgttatctgtatgggggacaatggaacaaa 284 |
| DB 213 | agaaaaagggggacaacacagtgagaatcgctctatccaaatggaacgcaatgggaaa 272 |
| QY 285 | tgaatcactaiaaagtaagaacctatactctttagcggagaataatcatctgtgtgt 344 |
| DB 273 | tgaatcctctggtgctgcgcgtgtgtgtctgcacaaacaaatggtttgacaaacgt 332 |
| QY 345 | tcttgaaagtatgtgtgacccggttatatcatctgtctgcataatcggtgtgtta 404 |
| DB 333 | gctgggaagctccacgagctacaggaagtatatacccgatgtattgataaagcgtg 392 |
| QY 405 | ttaattgattgaatagagaagtgctttaatttggaaagagaatataccgctattaat 464 |
| DB 393 | ttaattgattgaatagagaagtgctttaatttggaaagagaatataccgctattaat 452 |
| QY 465 | tgcgaatgaatggtttgtgtctgtggaaaggagatgtgttgctgaacgggtataaaga 544 |
| DB 453 | tgcgaatgaatggtttgtgtctgtggaaaggagatgtgttgctgaacgggtataaaga 512 |
| QY 525 | aatcccgcatgtcgtaaacgcgcttaaacacatactcgatgtagatgc 575 |
| DB 513 | gatcccgcatgtcgtaaacgcgcttaaacacatactcgatgtagatgc 563 |

RESULT 11
 ID 245343
 AC 245343;
 XX 27-MAR-2000 (first entry)
 XX DNA encoding a partial *Bacillus* sp. mannanase enzyme.
 XX

Search completed: December 20, 2000, 09:58:32
 Job time: 30586 sec

```

FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence
FT      misc-feature      2881..2940
FT      /tag= b
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
FT      misc-feature      4681..4740
FT      /tag= c
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
FT      EP786519-AZ.
FT      30-JUL-1997.
FT      07-JAN-1997; 97EP-0100117.
FT      05-JAN-1996; 96US-0009861.
FT      (HUMAN-) HUMAN GENOME SCI INC.
FT      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
FT      Rosen CA;
FT      WPI; 1997-374922/35.
FT      Polynucleotide(s) and proteins derived from Staphylococcus aureus
FT      stored on computer readable medium and used in the production of
FT      anti-S.aureus vaccines
FT      Claim 1: Page 1089-1092; 3271pp; English.
XX      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX      of the invention. The DNA sequences are recorded on a computer readable
XX      medium, preferably selected from a floppy or hard disk, random access
XX      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX      the S.aureus DNA sequences allows putative functions to be assigned so
XX      that protein-encoding or regulatory regions of commercial, therapeutic or
XX      industrial importance can be obtained. Specifically, sequences which are
XX      likely to encode antigens have been identified and these polypeptides can
XX      be used in a vaccine composition against S.aureus infection. The
XX      polypeptides can also be used in a kit for the immunodetection of
XX      S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX      skin and surgical wound infections, scalded skin syndrome, toxic shock
XX      syndrome, etc. Organisms transformed with the DNA sequences can be used
XX      for recombinant production of the polypeptides. The new DNA sequences
XX      (and their fragments) are useful as primers or probes for isolating
XX      homologues of any of the S.aureus DNA sequences contained on the
XX      computer readable medium.
XX      Sequence 4826 BP: 1645 A; 707 C; 893 G; 1400 T; 181 other;
SQ
  
```

Query Match 2.8%; Score 41.8; DB 18; Length 4826;
 Best Local Similarity 55.0%; Pred. No. 0.077;
 Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```

OY      896 attggcctggaataacttaacgtcttggaataacatagtgatgtcatatgtt 955
OY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      348 ATGGTGCTGGAAATTCCTTAGGTGATTCATGAATTAATTCCTCATGTAT 289
OY      956 taagagaactcgaagatgaacacgcttttacagtgagagatcgtatgaggaact 1015
OY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      288 TAATGGCTCTTATATATATATCGCTATTTAGGTGATGACTGTGTGCAACTT 229
OY      1016 ctccgacaactcttatgatttgaagt 1044
OY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      228 ATCAAGCAACAGGATTTGATTTAAAGT 200
  
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 01:23:33 ; Search time 75.74 Seconds
(without alignments)
2935.361 Million cell updates/sec

Title: US-09-339-159-1

Perfect score: 1470

Sequence: 1 ttgataatggttttaaaa.....ttgataatggtatgtagaa 1470

Scoring table: IDENTITY-NUC

Gap 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents-NA:*

1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/6_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/PCrUS_COMB.seq:*
7: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| C 1 | 48.6 | 3.3 | 7218 | 1 | US-08-232-463-14 |
| C 2 | 33.6 | 2.3 | 51259 | 5 | US-08-781-891-209 |
| C 3 | 33.4 | 2.3 | 1365 | 4 | US-08-870-827-4 |
| C 4 | 33.4 | 2.3 | 2408 | 4 | US-08-870-827-5 |
| C 5 | 33.2 | 2.3 | 3138 | 1 | US-07-867-106-4 |
| C 6 | 33.2 | 2.2 | 782 | 1 | US-08-261-825-1 |
| C 7 | 33.2 | 2.2 | 782 | 3 | US-08-719-124-1 |
| C 8 | 33.2 | 2.2 | 782 | 6 | PCR-US95-07748-1 |
| C 9 | 33.2 | 2.2 | 782 | 6 | PCR-US95-07748-1 |
| C 10 | 33.2 | 2.2 | 1800 | 5 | US-09-039-773A-1 |
| C 11 | 32.4 | 2.2 | 387 | 1 | US-08-222-177A-14 |
| C 12 | 32.4 | 2.2 | 5467 | 3 | US-07-745-206A-12 |
| C 13 | 32.4 | 2.2 | 3030 | 4 | US-08-311-363-12 |
| C 14 | 32.2 | 2.2 | 3030 | 4 | US-08-680-326-24 |
| C 15 | 32.2 | 2.2 | 8133 | 2 | US-08-480-604A-5 |
| C 16 | 32.2 | 2.2 | 8133 | 4 | US-08-405-496A-5 |
| C 17 | 32.2 | 2.2 | 4339 | 4 | US-08-566-398-23 |
| C 18 | 32.2 | 2.2 | 4365 | 6 | PCR-US91-08525-21 |
| C 19 | 32.2 | 2.2 | 4365 | 6 | PCR-US91-08525-25 |
| C 20 | 32.2 | 2.2 | 4365 | 6 | PCR-US93-04384-1 |
| C 21 | 32.2 | 2.2 | 4365 | 6 | PCR-US93-04384-7 |
| C 22 | 32.2 | 2.2 | 4365 | 6 | PCR-US93-04384-15 |
| C 23 | 32.2 | 2.2 | 6090 | 3 | US-08-566-398-35 |
| C 24 | 32.2 | 2.2 | 6144 | 3 | US-08-566-398-32 |
| C 25 | 31.8 | 2.2 | 7366 | 7 | 5169760-3 |
| C 26 | 31.8 | 2.2 | 7852 | 5 | US-08-836-022A-2 |

| | | | | | | |
|------|------|-----|-------|---|-------------------|-------------------|
| 27 | 31.8 | 2.2 | 7897 | 5 | US-08-836-022A-1 | Sequence 1, Appl |
| C 28 | 31.8 | 2.2 | 8809 | 2 | US-08-462-014-1 | Sequence 1, Appl |
| C 29 | 31.8 | 2.2 | 10398 | 3 | US-08-331-384-1 | Sequence 1, Appl |
| C 30 | 31.8 | 2.2 | 10398 | 3 | US-08-708-188-1 | Sequence 1, Appl |
| C 31 | 31.8 | 2.2 | 10398 | 3 | US-08-836-087-1 | Sequence 1, Appl |
| C 32 | 31.8 | 2.2 | 19182 | 4 | US-08-836-087-1 | Sequence 1, Appl |
| C 33 | 31.8 | 2.2 | 19182 | 4 | US-08-944-916-11 | Sequence 1, Appl |
| C 34 | 31.6 | 2.1 | 40328 | 5 | US-08-742-185-102 | Sequence 102, App |
| C 35 | 31.6 | 2.1 | 43795 | 5 | US-08-742-185-101 | Sequence 101, App |
| C 36 | 31.4 | 2.1 | 8855 | 3 | US-08-542-003-1 | Sequence 1, Appl |
| C 37 | 31.4 | 2.1 | 8855 | 3 | US-08-322-760A-1 | Sequence 1, Appl |
| C 38 | 31.4 | 2.1 | 1453 | 4 | US-08-252-995D-1 | Sequence 1, Appl |
| C 39 | 31.4 | 2.1 | 1453 | 4 | US-08-834-108-1 | Sequence 1, Appl |
| C 40 | 31.4 | 2.1 | 1600 | 1 | US-08-252-995D-5 | Sequence 5, Appl |
| C 41 | 31.4 | 2.1 | 1600 | 4 | US-08-834-108-5 | Sequence 5, Appl |
| C 42 | 31.4 | 2.1 | 2882 | 1 | US-08-424-567-1 | Sequence 1, Appl |
| C 43 | 31.4 | 2.1 | 2882 | 3 | US-08-711-928-1 | Sequence 1, Appl |
| C 44 | 31.4 | 2.1 | 3447 | 1 | US-08-252-995D-3 | Sequence 3, Appl |
| C 45 | 31.4 | 2.1 | 3447 | 4 | US-08-834-108-3 | Sequence 3, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELE: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14

GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,825
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-261-825-1
Query Match
Best Local Similarity 57.1%; Score 33; DB 1; Length 782;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1365 aaatgtccaaatcttctcagaaggaattggagtcagttccaatcagcgagtaga 1424
DB 35 AAATTACCAAAATCTTAATCAACATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1425 tagtagtgacaaacatcgattatattgataaagtgttaga 1469
DB 95 TAGCTCTGTCTAATCTTAATGATTGCTAATAGAAATGATA 139
RESULT 7
US-08-719-124-1
Sequence 1, Application US/08719124
GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,124
FILING DATE: 24-SEP-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,825
FILING DATE: 17-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-719-124-1
Query Match
Best Local Similarity 57.1%; Score 33; DB 3; Length 782;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 1365 aaatgtccaaatcttctcagaaggaattggagtcagttccaatcagcgagtaga 1424
DB 35 AAATTACCAAAATCTTAATCAACATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1425 tagtagtgacaaacatcgattatattgataaagtgttaga 1469
DB 95 TAGCTCTGTCTAATCTTAATGATTGCTAATAGAAATGATA 139
RESULT 8
PCT-US95-07748-1
Sequence 1, Application PC/TUS9507748
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

Query Match Similarity      2.2% Score 33 DB 6 Length 782;
Best Local similarity       57.1%; Pred. No. 2.6;
Matches 60; Conservative   0; Mismatches 45; Indels 0; Gaps 0;

QY 1365 aaatgccaagaattcttcacagtaagggaattggagttcagttcccaalcgcagtgta 1424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35 AAATACCAAAATCTTAATCAACATGCTTATTGGCGAACCAATAAGACTTTGT 94
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1425 tagtagtggacaacaacatcgatttatatgatatactgtatgtgaca 1469
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 TAGTGTTGTCTAATTCTTAATTAGTAGTGTCTAATAAGAATGATA 139
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT          9
PCT-US95-07748A-1
: Sequence 1, Application PC/TUS9507748A
: GENERAL INFORMATION:
: APPLICANT: The Regents of the University of California
: TITLE OF INVENTION: CLONED BORRELIA burgdorferi VIRULENCE
: TITLE OF INVENTION: PROTEIN
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: California
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/07748A
: FILING DATE: 16-JUN-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: HAITE, Ph.D., LISA A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07419/013WO1 (PD3516)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 678-5070
: TELEFAX: (619) 678-5099
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:

```

| Query Match | 2.2% | Score 33 | DB 6 | Length 782 |
|-----------------------|-------------------------------------------------------------|---------------|----------|------------|
| Best Local Similarity | 57.18 | Pred. No. 2.6 | | |
| Matches 60 | Conservative 0 | Mismatches 45 | Indels 0 | Gaps 0 |
| QY 1365 | aatgctcaaatcttcctcaatgaaggaattcgaggttcagttccatcagcgagtga | 1424 | | |
| | | | | |
| DB 35 | AAATTACCAAATCTTATTCACAAATGCTTAAATTTGCCAACCCAAATTAAGCTATTTGT | 94 | | |
| QY 1425 | tagtagtcagacacacatgcagattatattggaatcagtcgatttaga | 1469 | | |
| | | | | |
| DB 95 | TAGCTTGTTCTTAATTTCTTATTAGTATGCTTAATTAAGATGATTA | 139 | | |

RESULT 10
 US-09-039-773A-1
 ; Sequence 1, Application US/09039773A
 Patent No. 6100388
 GENERAL INFORMATION:
 APPLICANT: Casas, Ivan
 APPLICANT: Jonsson, Hans
 APPLICANT: Mvilstam, Bo
 APPLICANT: Roos, Stefan
 TITLE OF INVENTION: Lactobacilli Harboring Aggregation and Mucin
 TITLE OF INVENTION: Binding Genes As Vaccine Delivery Vehicles
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Standley & Gilcrest
 STREET: 495 Metro Place South, Suite 210
 City: Dublin
 STATE: Ohio
 COUNTRY: US
 ZIP: 43017
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44MB storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS Version 6.22
 SOFTWARE: Microsoft Word Version 6.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/039,773A
 FILING DATE: 16-MAR-1998
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA: No. 6100388 applicable
 ATTORNEY/AGENT INFORMATION:
 NAME: Donald O. Nickey
 REGISTRATION NUMBER: 29,092
 REFERENCE/DOCKET NUMBER: 1229-005
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (614) 792-5555
 TELEFAX: (614) 792-5536
 TELEX: No. 6100388 applicable
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1800 base pairs
 TYPE: Nucleic acid
 STRANDEDNESS: Double
 TOPOLOGY: Circular
 MOLECULE TYPE: Genomic DNA
 DESCRIPTION: Genomic DNA sequence and deduced amino
 acid sequence of bacterial aggregation
 protein

| | | | |
|----|-----|------------------------------------------------------------|-----|
| OY | 17 | aaaaatttttctatacaatcatcacttagtcagtcatactcggtgcttcag | 76 |
| | | | |
| Db | 113 | AAACAATTGTTTAAAACTTGTAATTCGATTTGGCAAAATTCCTCATCCTTTCTCG | 172 |
| | | | |
| OY | 77 | gaactctacgcgcataagccaattccggatttatatgaagcgtcacacctatacga | 136 |
| | | | |
| Db | 173 | TAGACACTAATCACGAATTCACAGGACTGTAAAACAGAACAATTTGAAGTATAAGCTT | 232 |
| | | | |
| OY | 137 | coaatgnaaccatttgttaatgaga | 162 |
| | | | |
| Db | 233 | ACCATTGAAAAGAACATTAAGGATA | 258 |

RESULT 12
US-07-745-206A-12/C
Sequence 12, Application US/07745206A
Patent No. 5429921
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Feldman, Daniel
TITLE OF INVENTION: Human Calcium Channel Compositions and
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitch, Even, Tabin & Flannery
STREET: 135 S. LaSalle
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/745,206A
FILING DATE: 19910815
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Feder, Scott B.
REFERENCE/DOCKET NUMBER: 51504
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-372-7842
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5467 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(144..3164, 3168..3245, 3249..3386, 3390..3392, 3396..3488, 3495..3539, 3543..3581, 3585..3587, 3591..3626, 3630..3689, 3693..3737, 3744..3746, 3750..4823, 4827..4841, 4845..5006, 5010..5096, 5100..5306, 5310..5366, 5370..5465)
US-07-745-206A-12

Query Match 2.2%; Score 32.4; DB 1; Length 5467;
Best Local Similarity 54.1%; Pred. No. 9.1;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 1275 tctgaagaacagcatgttatacatgtactctggaagcttctgtgcgattaacggttc 1334
DB 4884 tctgaagaacagcatgttatacatgtactctggaagcttctgtgcgattaacggttc 4825
QY 1335 atctggaagaacagcatctctgattatcaaatgtccaaatcttccaaagtaaggga 1394
DB 4824 atcttggcgcaaatccgaatctctgttactaaatattcagtaattactccaaacagtgga 4765
QY 1395 aa 1396
DB 4764 CA 4763

APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McCue, Ann
APPLICANT: Brenner, Robert
TITLE OF INVENTION: Human Calcium Channel Compositions and
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 166C Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,363
FILING DATE:
Prior Application Data:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-51506
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)238-0999
TELEFAX: (619)238-0062
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5467 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(144..3164, 3168..3245, 3249..3386, 3390..3392, 3396..3488, 3495..3539, 3543..3581, 3585..3587, 3591..3626, 3630..3689, 3693..3737, 3744..3746, 3750..4823, 4827..4841, 4845..5006, 5010..5096, 5100..5306, 5310..5366, 5370..5465)
US-08-311-363-12

Query Match 2.2%; Score 32.4; DB 3; Length 5467;
Best Local Similarity 54.1%; Pred. No. 9.1;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 1275 tctgaagaacagcatgttatacatgtactctggaagcttctgtgcgattaacggttc 1334
DB 4884 tctgaagaacagcatgttatacatgtactctggaagcttctgtgcgattaacggttc 4825
QY 1335 atctggaagaacagcatctctgattatcaaatgtccaaatcttccaaagtaaggga 1394
DB 4824 atcttggcgcaaatccgaatctctgttactaaatattcagtaattactccaaacagtgga 4765
QY 1395 aa 1396
DB 4764 CA 4763

RESULT 13
US-08-311-363-12/C
Sequence 12, Application US/08311363
Patent No. 5876958
GENERAL INFORMATION:

RESULT 14
US-08-680-326-24
Sequence 24, Application US/08680326
Patent No. 5925733
GENERAL INFORMATION:

```

Query Match Similarity      2.2%; Score 32.2; DB 4; Length 3030;
Best Local Similarity      53.6%; Pred. No. 8;
Matches    67; Conservative    0; Mismatches   58; Indels    0; Gaps    0.

Oy      51 agctagctcattctggttcgttcaggaaactctcaagctaatagcgaatccggattta 110
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      963 AACTGGAATAATGCATTATTCTTTAGGAACGCCCTCAGCTGTGGAAAAAATCGAACTTAA 1022
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy     111 tctaacggttacacactctacagtcatgccaatggaaccatttgtaatatgagaagattaa 170
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     1023 TCAAGTCCCTACTGAAATAGACATGCTGCATGGGTTTTTTTCATTATATAGACACTTFAA 1082
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy      171 ccaatg 175
          |||||
Db     1083 TGCTTG 1087

RESULT 15
US-08-480-604A-5
; Sequence 5, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BROCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP

```

```

Query Match          2.2%; Score 32.2; DB 2; Length 8133;
Best Local Similarity 49.1%; Pred. NO. 12;
Matches      85; Conservative      0; Mismatches 88; Indels      0; Gaps
QY      2 tgaataatgggttttaaaaaaatttttcataacgtctatcattactatgctagctcta 61
Db      5333 TGAGTATGATTTTAAAGTATTATAAAAAACTTCATTTAGAGATATATTAAGAGTAATTTTA 5392
QY      62 ttctgttcggttcaggaaactctcacgctaacycaaatccggatttctgttaacgcgcta 121
Db      5393 TATCATTTTATATCTCTAAAATGATATTAGATAGAGATCATTTTAGAGTTTAAAATATAGATA 5452
QY      122 ccacactatagcatgcacaatggaaacccacttgtaatgagaggatataccat 174
Db      5453 ATTTAAACTTATTCATATGATGAAGATAGTAATTAATTGTTAAAGGATTAATCAAT 5505

Search completed: December 20, 2000, 09:56:10
Job time: 30757 sec.

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 19, 2000, 23:01:31 ; Search time 1967.85 Seconds
(without alignments)
4618.607 Million cell updates/sec

Title: US-09-339-159-1
Perfect score: 1470
Sequence: 1 ttgaataatggttttaaaaa.....ttgataatgtgattgtagaa 1470

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estda:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

117: gb_gsa18:*
 118: gb_gsa19:*
 119: em_gsa13:*
 120: gb_gsa20:*
 121: gb_gsa21:*
 122: gb_gsa22:*
 123: gb_gsa23:*
 124: gb_gsa24:*
 125: em_gsa14:*
 126: em_gsa15:*
 127: em_gsa16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|-----|----------|---------------------|
| 1 | 46.2 | 3.1 | 1101 | 121 | CNS000D1 | AL065414 Drosophila |
| 2 | 46.2 | 3.1 | 1101 | 121 | CNS00396 | AL065921 Drosophila |
| 3 | 43.4 | 3.0 | 1101 | 121 | CNS00172 | AL078714 Drosophila |
| 4 | 43.4 | 3.0 | 1101 | 121 | CNS0182P | AL108811 Drosophila |
| 5 | 42.8 | 2.9 | 1101 | 121 | CNS00E81 | AL069797 Drosophila |
| 6 | 42.2 | 2.9 | 1101 | 121 | CNS01412 | AL104180 Drosophila |
| 7 | 41.8 | 2.8 | 471 | 114 | AZ028169 | AZ028169 RPII-23-3 |
| 8 | 41.2 | 2.8 | 350 | 36 | C90049 | C90049 CS0049 Dict |
| 9 | 40.4 | 2.7 | 1101 | 121 | CNS00B3U | AL055389 Drosophila |
| 10 | 40.2 | 2.7 | 303 | 30 | BB388799 | BB388799 BB388799 |
| 11 | 40.2 | 2.7 | 631 | 15 | AU061548 | AU061548 AU061548 |
| 12 | 39.4 | 2.7 | 552 | 96 | AQ332958 | AQ332958 HS_5001_A |
| 13 | 39.2 | 2.7 | 520 | 100 | AQ621653 | AQ621653 HS_3089_B |
| 14 | 39.2 | 2.7 | 1204 | 121 | CNS016E2 | AL106628 Drosophila |
| 15 | 39.2 | 2.7 | 750 | 115 | AZ138670 | AZ138670 SP_0170_A |
| 16 | 38.8 | 2.6 | 459 | 21 | AM333544 | AM333544 S23A10 AG |
| 17 | 38.8 | 2.6 | 515 | 21 | AM333565 | AM333565 S4ED3 AGS |
| 18 | 38.8 | 2.6 | 545 | 21 | AM334819 | AM334819 S39E11 AG |
| 19 | 38.8 | 2.6 | 827 | 121 | CNS00E03 | AL069854 Drosophila |
| 20 | 38.8 | 2.6 | 1000 | 121 | CNS00C0C | AL055447 Drosophila |
| 21 | 38.4 | 2.6 | 1101 | 121 | CNS0106X | AL098595 Drosophila |
| 22 | 38.4 | 2.6 | 250 | 15 | AU060486 | AU060486 AU060486 |
| 23 | 38.4 | 2.6 | 295 | 15 | AV098721 | AV098721 AV098721 |
| 24 | 37.8 | 2.6 | 306 | 27 | BB137465 | BB137465 BB137465 |
| 25 | 37.8 | 2.6 | 309 | 25 | AM953387 | AM953387 EST365457 |
| 26 | 37.8 | 2.6 | 510 | 25 | AM943977 | AM943977 LP01629_3 |
| 27 | 37.6 | 2.6 | 488 | 15 | AU050493 | AU050493 AU050493 |
| 28 | 37.6 | 2.6 | 524 | 90 | AQ204415 | AQ204415 HS_3218_A |
| 29 | 37.6 | 2.6 | 674 | 124 | CNS0543D | AL320242 Tetradodon |
| 30 | 37.6 | 2.6 | 933 | 36 | BE349287 | BE349287 601077980 |
| 31 | 37.6 | 2.6 | 1101 | 121 | CNS00YWL | AL096927 Drosophila |
| 32 | 37.4 | 2.5 | 276 | 18 | AV324502 | AV324502 AV324502 |
| 33 | 37.4 | 2.5 | 486 | 4 | AA429143 | AA429143 zw51d01_r |
| 34 | 37.4 | 2.5 | 510 | 97 | AQ413085 | AQ413085 RPII-11-1 |
| 35 | 37.4 | 2.5 | 545 | 98 | AQ534365 | AQ534365 RPII-11-3 |
| 36 | 37.4 | 2.5 | 601 | 96 | AQ375742 | AQ375742 SP_1031_B |
| 37 | 37.4 | 2.5 | 717 | 116 | AZ186394 | AZ186394 SP_1031_B |
| 38 | 37.4 | 2.5 | 1101 | 121 | CNS017XD | AL108811 Drosophila |
| 39 | 37.2 | 2.5 | 411 | 99 | AQ593899 | AQ593899 HS_2085_A |
| 40 | 37.2 | 2.5 | 586 | 24 | AM776280 | AM776280 EST335345 |
| 41 | 37.2 | 2.5 | 618 | 91 | AQ271069 | AQ271069 nbd0017N |
| 42 | 37.2 | 2.5 | 1080 | 121 | CNS00E8P | AL069949 Drosophila |
| 43 | 37.2 | 2.5 | 849 | 108 | AQ740244 | AQ740244 HS_5505_A |
| 44 | 37 | 2.5 | 938 | 121 | CNS006T1 | AL065906 Drosophila |
| 45 | 36.8 | 2.5 | 521 | 115 | AZ129098 | AZ129098 OSUNdb009 |

ALIGNMENTS

RESULT 1
 CNS000D1

| | | | | | |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-------|-------|-------------|
| LOCUS | CNS000D1 | 1101 bp | DNA | GSS | 03-JUN-1999 |
| DEFINITION | Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR01J16 of RPII-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | | | |
| ACCESSION | AL065414 | | | | |
| VERSION | AL065414.1 | GI:4938827 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | fruit fly. | | | | |
| ORGANISM | Drosophila melanogaster | | | | |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | | |
| REFERENCE | Genoscope. | | | | |
| AUTHORS | Submitted (02-JUN-1999) Genoscope - Centre National de Sequenc | | | | |
| TITLE | BP 191 91006 EVRI cedex - FRANCE (E-mail : seque@genoscope.cns.fr | | | | |
| JOURNAL | - Web : www.genoscope.cns.fr) | | | | |
| COMMENT | Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammocser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPII-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..1101 | | | | |
| | /organism="Drosophila melanogaster" | | | | |
| | /db_xref="taxon:7227" | | | | |
| | /clone_lib="RPII-98" | | | | |
| | /clone="BACR01J16" | | | | |
| | /note="end : TET3" | | | | |
| BASE COUNT | 280 a | 104 c | 123 g | 211 t | 383 others |
| ORIGIN | | | | | |
| Query Match | 3.1% Score 46.2; DB 121; Length 1101; | | | | |
| Best Local Similarity | 14.5% Pred. No. 0.069; | | | | |
| Matches | 72; Conservative 222; Mismatches 202; Indels 2; Gaps 1; | | | | |
| QY | 644 | aacgaatacatgttgcattcatatgataatgataatgacagtgatgacatgacgaag | 703 | | |
| DB | 594 | AAAAMAGRGATNDKBAKMDKAAADAKAKAAAHAGTWTTKDKKKKKKKKKKKAKR | 653 | | |
| QY | 704 | ttgtactaatatgacccaggttcttaatacaagactgcattagtcattggtgaattg | 763 | | |
| DB | 654 | AATDGTMTATTTWTKTKAKDGAKKAKKADRTKAMKACAGARATAKARAKKRWMD | 713 | | |
| QY | 764 | gacacccatacaaatggtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacaa | 823 | | |
| DB | 714 | KDTKKKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK | 773 | | |
| QY | 824 | gaggaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt | 883 | | |
| DB | 774 | GARDKDKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK | 833 | | |
| QY | 884 | acccttgaaatgat--tggctgtggaataacacccatgacgttggtggaataacatagt | 941 | | |
| DB | 834 | ADKKKAGKKKKWTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK | 893 | | |
| QY | 942 | tggtccatagtgttaagagaacttcgaactgaacacgttttaacaggtgtggaagac | 1001 | | |
| DB | 894 | WDODDTAKDDMTTKKDDMTWATAGATAGWDGDKAKGDKRGTDWTKTKTKTKTK | 953 | | |
| QY | 1002 | tgatgaggaacttctccgacaactcttatgatttgaagtagatgcaagatgac | 1061 | | |

| DB | 954 | KTAGACGCGKDKKDKNDKRDATGCKDWTWBDTDTOTAKRRRRRRRRRAGGGRDKD | 1013 |
|-----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|----------------------|
| Oy | 1062 | tggaagtgcttgagcggaggtccttggtctgtgacagaggtcttctcaaggagta | 1121 |
| Db | 1014 | KKRRRAARRRRKKKKTKDKDDGGRWRAAKAKADAAKADADAADAAAMAAKAKAKD | 1073 |
| Oy | 1122 | ttctttaaaggcggatct | 1139 |
| Db | 1074 | DKAKKADADDDKDKKAK | 1091 |
| RESULT 2 | | | |
| CNS00396/C | | | |
| LOCUS | | | |
| DEFINITION | Drosophila melanogaster genome survey sequence Tm3 end of BAC # BACR08R10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | |
| ACCESSION | AL063921 | | |
| VERSION | AL063921.1 | | |
| KEYWORDS | GSS. | | |
| SOURCE | fruit fly. | | |
| ORGANISM | Drosophila melanogaster | | |
| REFERENCE | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | |
| AUTHORS | 1 (bases 1 to 1101) | | |
| TITLE | Genoscope. | | |
| JOURNAL | Direct Submission | | |
| COMMENT | Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseogawa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . Location/Qualifiers 1. 1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_11b="RPCI-98" /clone="BACR08R10" /note="end : Tm3" BASE COUNT 201 a 64 c 131 g 202 t 503 others ORIGIN | | |
| FEATURES | | | |
| SOURCE | | | |
| Query Match | 3.1% | Score 46.2; | DB 121; Length 1101; |
| Best Local Similarity | 14.5% | Pred. No. 0.069; | |
| Matches 65; Conservative 215; Mismatches 164; Indels 3; Gaps 1; | | | |
| Oy | 773 | atacaaatgtaagctgatagaagaacgatatgaagctatctgaaacaagaagagt | 832 |
| Db | 1026 | WWWWATAKTADTAWTWKTAWRADWAGNDRGAGKRDRAATADGAGRGDGRKKRDKD | 967 |
| Oy | 833 | gtgtgtgtcgctgtgtcatgaaaggaagcgcccaagatggagagattatagaccttga | 892 |
| Db | 966 | RKDGDGDKKGGKKRAAKAAKAWATWMDWMDWMDKMKWGDARRADDGAGDKDDDK | 907 |
| Oy | 893 | atgatggctgtaaatcaacttaacagctctgggggaataataatgatgatgttcata | 952 |
| Db | 906 | GADADDUDTGTDDDDKDKWDAKGTGWATWAAATDWWMGWADADWWTDAAD | 847 |

```

QY 953 gtttaagagaacttcga---gattaagaccgctttttacaggctgagagatcaggag 1009
Db 846 DWWADDKRDWAMWKKDDAMAMGARFARDRDWMDRACKRGAGAKRRDRKRAADKRDADAADR 787
QY 1010 gaactcttcgcgaacctctttatgatttgaagagtagatgaagagatgaggaagta 1069
Db 786 DDAATWTTWTTTTRDTDDWKKMTDTWTRWADRTWDRDDDDDRBAGTAGAKWRRTWRK 727
QY 1070 gctttaagcggagagtccttggtcgtgacacagatgctctctaaggaagatctttta 1129
Db 726 RMRKRDFTWTDADADTADRDARRRGDDADAGKCKTGKRRRRRRRAWRTDAMWADA 667
QY 1130 aagcgagatattcattgctgcataacttcacacattcatctatgatttcatcaataagc 1189
Db 666 AAWTTTDDTDDWCKDRDRRRKRGARRRRRTTAAGANDWTKAMDMWAKMKTRADRWDX 607
QY 1190 cttacagcgaagatagtagatacaag 1216
Db 606 AADTWDARKADRDWAKARAWRARDR 580

RESULT 3
CNS00LT2 1101 bp DNA GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACRA48P19 of RPCT-98 library from Drosophila melanogaster (fruit
VERSION AL078714.1 GI:5102004
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Archipoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqetf@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the Blue
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_11b="RPCT-98"
/clone="BACRA48P19"
/note="end : TET3"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

ORIGIN
Query Match 3.0%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 14.9%; Pred. No. 0.4; Indels 0; Gaps 0;
Matches 48; Conservative 140; Mismatches 133;

315 tttaagcgaagatacatcttggtcgtcttcttgaaaltcatagatgctaccggtatga 374

```


| | | | |
|----|------|----------------------------------------------------------------|------|
| Db | 752 | MMWTDAITWADTWMKAATYTDLAKRAAGRRKKDAKTAARDGRRARTRRRRMAAGRRARG | 811 |
| QY | 788 | tcgatgaagcgaacgattatgagctatcttgcacaagaagagtttggttgcgtggt | 847 |
| Db | 812 | ARRRARRRADDRCMDVMWMAAAAAAAAAAAATWTRDRSMWMDMDWDTRMDDTTAAAMD | 871 |
| QY | 848 | catgaaaggaagcgcgcgaagctggagatcttgaacttcgaaatggttgctgaa | 907 |
| Db | 872 | DARARARRRRRRRRRRARRARRRRAADDTDKDRNADATTKDKTTWTTDDDDWDNAK | 931 |
| QY | 908 | ataacctacagcttggggaatacataatgtagatggtccatagtttgaagaagaact | 967 |
| Db | 932 | RDRWMAAKADGAWKMRNDPRANDMAATANKDDGKKDKWGRGKRGGKDKRMPKXGTGKD | 991 |
| QY | 968 | cgagattaaagcacccgtttttacagctgtagagatcctgatggaagaacttcgcgaacatc | 1027 |
| Db | 992 | DDDDMDKTTWTDWMMWWTRTKMDMMMDGCGRGKTRKRWGAWMADAWARDDTDCDKDRT | 1051 |
| QY | 1028 | tttctgatttgaagtagatgcgaagctggacgcgaagtagtgccttg | 1076 |
| Db | 1052 | ADKRKRKTTDTTKDGDGDKRDKRRKDKCDDTKDKATWTTDDDDAD | 1100 |

RESULT 5
 CINSO0EST/C
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

CINSO0EST 1101 bp DNA GSS 04-JUN-1999
 Drosophila melanogaster genome survey sequence TET3 end of BAC:
 BACR39L15 of RP11-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL069797
 AL069797.1 GI:4949738
 GSS
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osogawa and
 Aaron Mamoeris at the Drosophila Laboratory in the Department of

FEATURES
source
1. .1101
location/qualifiers

| | | | | | |
|------------|-------|-------|------|-------|------------|
| BASE COUNT | 174 a | 179 c | 72 g | 207 t | 469 others |
| ORIGIN | | | | | |

| | | | | |
|-----------------------|-------------------|-----------------|-----------|--------------|
| Query Match | 2.9%; | Score 42.8; | DB 121; | Length 1101; |
| Best Local Similarity | 19.7%; | Pred. No. 0.58; | | |
| Matches 122; | Conservative 196; | Mismatches 300; | Indels 0; | Gaps 0 |

| QY | 5 | ataaagcgtttcaaaaataattttctatacaactatcatcatatgaagtaagctcattc | 64 |
|-------------|---------------------------------------------------------------------|---------------------------------------------------------------|------|
| Db | 1101 | AKAMWKKWADMDADATATTKRWATDARATAAKADGAMGRRTGAADRGNWTGRTATTA | 1042 |
| QY | 65 | tgttcgttcaggaaactctacaagctaaatgcgaattccgaatttatgtlaagcggatcca | 124 |
| Db | 1041 | WTTADTTGGRGAKRRGRGKGRNDTWTAKWGTATFRRTGTTWDDMWKKTFATATARA | 982 |
| QY | 125 | ctctatacgaatgccaatggaaccatttgttatgaaggggttaaccaatggcgcgct | 184 |
| Db | 961 | GTAKGAAGAKAGTAATWADAKADATATKDKDWATSDTFRGMAAGRTWGMGTTDDMKACT | 922 |
| QY | 185 | gataaagaacccaagcaactaactgcaattggaaggatgccaataccggtgctataacg | 244 |
| Db | 921 | CAKTTTMMKMGATTCGAGAGAGAGAGCGMGRMCMWDAGTRMGWAKADDKKMDTGWK | 862 |
| QY | 245 | tccgagatgtgtatcgtatggtgggacaatgagacaagaatgacatccatacagtaagaa | 304 |
| Db | 861 | TGTDWCGTTTKRMAATTTWKTFRMTARATATWATADADDMAKRRKGMATTAARFTTKAR | 802 |
| QY | 305 | acctatctcttagcggaaagtaatccttggctgcgcttccttgaagtcagtcagtcga | 364 |
| Db | 801 | AMTARATATTAAGATAGADADADKGTMMGAAMWGTTRADAKAKTRTADATTTAAKAT | 742 |
| QY | 365 | ccggtatgaatccatctgcttcgccaactcgctgctgattatcgaattgaatgaaga | 424 |
| Db | 741 | GTGKKKGAAGATRTMDWDMGRTMDKAAARARACMAARTRRGMAARADRMALAAATATWR | 682 |
| QY | 425 | gtgccttaatcggaaagaaataccgcgtcatatgaatgacgaatgaagtggttgct | 484 |
| Db | 681 | TGTWATATMTATRTMTGGTDTTRKTRARAGTKTKKMAKAKTTDDATATDNRKTTTGWCD | 622 |
| QY | 485 | ctgtggaaggggagatcctgtgsgctgcgggtataaacaagcaatcccgatgctgaacg | 544 |
| Db | 621 | GDKGWGRGKGGKGGKGTGAATRAARVGGKATKADMDWDTTWTWKMTTKDRMTTKT | 562 |
| QY | 545 | ccggtcgaaccatcccttgaatgtagatgctgcgggtgaggaacattccacaatga | 604 |
| Db | 561 | GTGRTTWTMTTWTARWAGAKARAVGTDMWDMARARACSSSSRRWAMGSRVTATATNA | 502 |
| QY | 605 | tctatgatattcgaagaag | 622 |
| Db | 501 | KRGWGTDTATCSARKAG | 484 |
| RESULT | 6 | | |
| CNS01412 | | | |
| LOCUS | | | |
| DEFINITION | CNS01412 1101 bp DNA GSS 26-JUL-1999 | | |
| DESCRIPTION | Drosophila melanogaster genome survey sequence sp6 end of BAC | | |
| ACCESSION | BA001119 of DrosBAC library from Drosophila melanogaster (fruit | | |
| VERSION | AL104180 | | |
| KEYWORDS | AL104180.1 GI:5615791 | | |
| SOURCE | GSS. | | |
| ORGANISM | fruit fly. | | |
| REFERENCE | Drosophila melanogaster | | |
| AUTHORS | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| JOURNAL | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; | | |
| COMMENT | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | |
| | 1 (bases 1 to 1101) | | |
| | Genoscope. | | |
| | Direct Submission | | |
| | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage | | |
| | BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr | | |
| | - Web : www.genoscope.cns.fr) | | |
| | Determination of this BAC-end sequence was carried out as part of a | | |
| | collaboration with the European Drosophila Genome Project (BDGP) - | | |
| | http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC | | |
| | library (Dros BAC) was made by Alain Billand at CPH (Centre | | |
| | d'Etude du Polymorphisme Humain) with funding provided by a MRC | | |
| | project grant. The DNA was prepared from embryos by Alain Bucheton | | |
| | and Genevieve Payan. It has been constructed in the vector | | |

| FEATURES | Location/Qualifiers |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Source | 1. 1101 |
| | /organism="Drosophila melanogaster" |
| | /plasmid="pBelobAC11" |
| | /db_xref="taxon:7227" |
| | /clone_lib="DrosBAC" |
| | /clone="BACN11119" |
| | /note="end : SP6" |
| BASE COUNT | 380 a 159 c 215 g 262 t 85 others |
| ORIGIN | |
| Query Match | 2.9%; Score 42.2; DB 121; Length 1101; |
| Best Local Similarity | 40.8%; Pred.No.0.85; Indels 1; GTC |
| Matches 125; Conservative | 30; Mismatches 150; |
| QY 1075 | agcgagagctcttgaggcgtgtgacagagtgctctcttaagaagcattctttaaacgc 1134 |
| DB 722 | AGGCCAGGCTTGWGTTTCAGMNAACCCCAATCGAGTGGGATGTGATAACAGCG 781 |
| QY 1135 | gatactcaattgttcgccaattcaacaactcttaccatgttatccaatacagctctta 1134 |
| DB 782 | GASAKGCTGTAG-CTTAAAGCTCGWATTTKCGATGCCGAGTGCACAAATAATCATCAG 842 |
| QY 1195 | cagcagaatctgtgatacaagctactgttaacatgtccaattgtgggaagtggtgtat 1134 |
| DB 841 | CTCGACGMAACMGAAAGAAAGMAACCAATWGTATAGTGTGAGGAAAGATGCAAAW 900 |
| QY 1255 | ggaatgactgcggtctcttatgtgaaacaagcagctgtatatacagtgactcgtgaac 1314 |
| DB 901 | TWATTTTTCGNTTYTARAGTCGTAAWMAATATATDGTGTTGCGACACACACAAAGAC 960 |
| QY 1315 | ttgtgcggaatacaggttcatctgtgaaacaagcgtatctctagattatcaaatgtccaa 1374 |
| DB 961 | TGGGWTTWAAACRKGAGACAGCTAGRAGASGCTTACTGWWCACTATTAGKCTCMA 1620 |
| QY 1375 | aatctt 1380 |
| DB 1021 | AMAGTT 1026 |
| RESULT 7 | |
| LOCUS | AZ028169/c |
| DEFINITION | AZ028169 471 bp DNA GSS 25-FEB-2000 |
| ACCESSION | RFCl-23-31608-TV RFCl-23 Mus musculus genomic clone RFCl-23-31608, |
| VERSION | AZ028169 |
| KEYWORDS | AZ028169.1 GI:7103553 |
| SOURCE | GSS. |
| ORGANISM | house mouse, |
| REFERENCE | Mus musculus |
| AUTHORS | Eukaryotic Metazoa: Chordata; Craniata; Vertebrata: Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 471) |
| | Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akiret ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and and Fraser,C.M. |
| TITLE | Mouse BAC End Sequences from Library RFCl-23 |
| COMMENT | Other_GSSs: RFCl-23-31608-TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RFCl-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: |

```

Matches      55; Conservative      85; Mismatches      100; Indels      0; Gaps      0.

```

Db 617 BGAGMAMKSKTMDTKMTBTBBYADMMWCATYMTTKYKYYAAADGTTTKKRAK 558
 Oy 335 tgggtgctgtcttgaagtcattgacgtaccgttatgattcattcgttcgtcgaac 394
 Db 557 GTMTAKACBSBYYTAAKMAATADMMTKKAMWKTTTAAATTKMDATATWTKTKTKTKA 498
 Oy 395 gtgcgtgattattggttggaagaagtcgttatttgaagaagaagaccgtca 454
 Db 497 AMCKTKTAKKKKKKKKTGTAAGAKKAKKAAAMKTAKTKTKKAKAAAMADATTTPAA 438
 Oy 455 ttattaatatcgcaatgattggttgcgtgcgggaagggagtcgttcggtcgcgggt 514
 Db 437 KAKKKAATKTAADGATGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 378

RESULT 10
 LOCUS BB388799/c
 DEFINITION Musculus full-length enriched, 0 day neonate cerebellum Mus musculus cDNA clone C230058N01 3' similar to U31668 Ratius norvegicus transcription factor E2F-5 mRNA, mRNA sequence.

ACCESSION BB388799
 VERSION BB388799
 KEYWORDS GI:9111642
 SOURCE EST.
 ORGANISM house mouse.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 303)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Haru, A., Hayasaka, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, U., Ishikawa, T., Itoh, M., Iizawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, K., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yamana, T., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 Genome Science Laboratory
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9033
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp,
 url:http://genome.rtc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagasaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermosensitization and thermocactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Iizawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 SOURCE Location/Qualifiers
 1..303 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="C230058N01"

/clone_1lb="RIKEN full-length enriched, 0 day neonate cerebellum"
 /tissue_type="cerebellum"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'] GAGGAGAGAGATCCAGACGCTTTTCTTTTCTTTTCTTTT 3'1]. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5'] GAGGAGAGATCTTCGAGTTAATTAAATTCCTCCCTCC 3'1]. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 70 a 77 c 43 g 113 t
 ORIGIN

Query Match 2.7%; Score 40.2; DB 30; Length 303;
 Best Local Similarity 54.4%; Pred. No. 2;
 Matches 81; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Oy 993 tggagatcatatgaggaatcttcgcacacatttatgatttgaagtagatgca 1052
 Db 253 TGGGACATCTTATATATCAACATCAACATCAACATCTTCCGCTACTCTAAGAAAGTGA 194

Oy 1053 aggatgacgtgaagtagcttgagcgaggtccttggtgacagagtggtctca 1112
 Db 193 AGGTGAGAAATTCACAACTGAAAGGGGTTAAGGGGCTGAACAGAAAGACTCCA 134

Oy 1113 aggaagtcattctttaaagcggatattc 1141
 Db 133 AAATCGGCTCACTTAACAGTGAATATCC 105

RESULT 11
 LOCUS AU061548/c
 DEFINITION AU061548 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium discoideum cDNA clone SLE608, mRNA sequence.

ACCESSION AU061548
 VERSION AU061548
 KEYWORDS GI:4882652
 SOURCE EST.
 ORGANISM Dictyostelium discoideum.
 Dictyostelium discoideum.
 Eukaryota; Dictyostelida; Dictyostelium.
 1 (bases 1 to 631)
 Yoshino, R., Morio, T. and Tanaka, Y.
 Developmental cDNA in Dictyostelium discoideum
 Unpublished (1997)
 Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
 Email: d402n@esakura.cc.tsukuba.ac.jp
 PROJECT = Dictyostelium discoideum cDNA project in Japan.
 Location/Qualifiers
 1..631 /organism="Dictyostelium discoideum"
 /strain="Ax4"
 /db_xref="taxon:44689"
 /clone_1lb="Dictyostelium discoideum SL (H.Urushihara)"
 /dev_stage="slug"

BASE COUNT 262 a 194 c 84 g 91 t

ORIGIN

| | | | | |
|-----------------------|--------------|---------------|----------------|------------|
| Query Match | 2.7% | Score 40.2 | DB 15 | Length 631 |
| Best Local Similarity | 47.8% | Pred. No. 2.5 | | |
| Matches 117 | Conservative | 0 | Mismatches 128 | Indels 0 |
| | | | | Gaps 0 |

| | | | |
|----|------|-----------------------------------------------------------------|------|
| QY | 893 | atgattggctggcggaaaataaccttcacagcttggggaaatacatatgaatggtccatctg | 952 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 430 | atgattgactgttggttttggaatcagcttggctgtgttcacacatttgatgatgtcaggtg | 371 |
| | | | |
| | | | |
| | | | |
| | | | |
| QY | 953 | gtttaagagaacttcgagattaagacccgttttcacagtgtagagatctgattggaa | 1012 |
| | | | |
| | | | |
| | | | |
| Db | 370 | gttggtcagcagatttcacgtcaggaacctctgtctctggtgcagctgatgttcagctg | 311 |
| | | | |
| | | | |
| | | | |
| | | | |
| QY | 1013 | cttctcgaacaccttattgattttgaaagttagtgcgaagatgcgaatgaaagtaact | 1072 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 310 | gttggtgttttaccagttgagagatgatgtgacctgcttcatttgatttgcctggttctt | 251 |
| | | | |
| | | | |
| | | | |
| | | | |
| QY | 1073 | tgaacggagagtcctctggcctgtgcacagagtcgtcttcctctaaaggaagtcactttaaag | 1132 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 250 | cgcgcagctgagatgttgacgtgcagaccttcatttgatttctgctgctgttcttccagcag | 191 |
| | | | |
| | | | |
| | | | |
| | | | |
| QY | 1133 | cgagat | 1137 |
| | | | |
| | | | |
| Db | 190 | ctgagt | 186 |

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| RESULT | 12 |
| A0332958/c | |
| LOCUS | |
| DEFINITION | |
| <p>A0332958 552 bp DNA GSS 06-MAR-1999</p> <p>HE_5001.A2_F09.T7 RFL11 Human Male BAC Library Homo sapiens</p> <p>genomic clone Plate=577 Col=18 Row=Y, DNA sequence.</p> | |

| RECORDS | SOURCE | ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL | MEDLINE | COMMENT |
|---------|--------|---------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|---------------------------------|---------------------------|------------------------------------------|
| 9331 | human. | <i>Homo sapiens</i> | 1 (bases 1 to 552) | Mahatras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L. | Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome | Proc. Natl. Acad. Sci. U. S. A. | 96 (17), 9739-9744 (1999) | Contact: Mahatras GG, Wallace JC, Hood L |

| | | | | | |
|------------|-------|-------|------|-------|----------|
| BASE COUNT | 177 a | 133 c | 67 g | 170 t | 5 others |
| ORIGIN | | | | | |

| | | | | |
|-----------------------|--------------|--------------|----------------|-------------|
| Query Match | 2.7% | Score 39.4; | DB 96; | Length 552; |
| Best Local Similarity | 53.7%; | Pred. No. 4; | | |
| Matches 79; | conservative | 0; | Mismatches 68; | Indels 0; |
| | | | | Gaps |

| | | | | |
|-----|----|-----|-------------------------------------------------------------|-----|
| 353 | QY | 234 | tacagtaagaacctctctcctcttaacggaataataccttgggttgctgctgaagt | 353 |
| 182 | DB | 182 | TAAAGTAAATATCTATTATTAAGTGAATAPAGAAATATATCTTGCCCTTGATGTAAGT | 122 |
| 354 | QY | 354 | tcatgatgcacgcggttatgattccattgcttcgcgaatgcgtctgttgaattgaat | 413 |
| 122 | DB | 122 | CCAAAGCTGCCAGGTTTGTGNTTTGATGTCAGTCACTCATTCATGTTCATGAATTTNGT | 63 |
| 414 | QY | 414 | tgaatggaagtgccttaattgnaa | 440 |
| 62 | DB | 62 | TGATGAGATAGTAAATCAATCGAG | 36 |

| | |
|------------|-------------------------------------------------------------------|
| RESULT | 13 |
| A0621653 | |
| LOCUS | 520 bp DNA |
| DEFINITION | HS_3089_B2_H12_MF CIT Approved Human Genomic Sperm Library D Homo |
| ACCESSION | sapiens genomic clone Plate=3089 Col=24 Row=P, DNA sequence. |
| VERSION | A0621653 |
| KEYWORDS | A0621653.1 GI:5084045 |
| SOURCE | GSS. |
| | human. |

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .520 |

| BASE COUNT | 146 a | 100 c | 129 g | 143 t | 2 others |
|------------|-------|-------|-------|-------|----------|
| ORIGIN | : | | | | |

| | | | | |
|-----------------------|-------|----------------|---------|----------------|
| Query Match | 2.7% | Score 39.2; | DB 100; | Length 520; |
| Best Local Similarity | 57.3% | Pred. No. 4.5; | | |
| Matches | 71; | Conservative | 0; | Mismatches 53; |
| | | | | Indels 0; |
| | | | | Gaps |

| QY | 313 | tccttagcggagataatcatttggtgctgctcttgaagtcacagaccgggtat | 372 |
|----|-----|---------------------------------------------------------|-----|
| | | | |
| Db | 211 | tccttttcgcaatcacagactctttatttttttttcggttaagtgccacaggttt | 270 |
| | | | |
| QY | 373 | gattcattgcttgcgtcaacgcgtgcttgattatgattgaatgagaagtgcctta | 432 |

| | | | |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|----------------------|
| Db | 271 | GTTTAATGATTGCCTTCAGGATTTGGTTCAGGATTGCACTAAAGAAGATGACATTTA | 330 |
| Oy | 433 | attg 436 | |
| Db | 331 | AATG 334 | |
| | | | |
| RESULT 14 | CNS016E2 | 1204 bp | DNA |
| LOCUS | CNS016E2/C | | GSS |
| DEFINITION | Drosophila melanogaster genome survey sequence T7 end of BAC BACH15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | |
| ACCESSION | AL106628 | | |
| VERSION | ALI06628.1 | | GI:5622852 |
| KEYWORDS | GSS. | | |
| SOURCE | fruit fly. | | |
| ORGANISM | Drosophila melanogaster | | |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| | Musculophila; Neoptera; Endopterygota; Diptera; Brachycera; | | |
| | Pteromorphia; Ephydroidea; Drosophilidae; Drosophila. | | |
| REFERENCE | 1 (bases 1 to 1204) | | |
| AUTHORS | Genoscope. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreifgenoscope.cns.fr | | |
| COMMENT | - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration of this BAC-end sequence was carried out as part of a http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphism Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11. | | |
| FEATURES | Location/Qualifiers | | |
| Source | 1..1204 | | |
| | /organism="Drosophila melanogaster" | | |
| | /plasmid="pBelOBAC11" | | |
| | /db_xref="taxon:7227" | | |
| | /clone_1lb="DrosBAC" | | |
| | /clone="BACH15A12" | | |
| | /note="end : 17" | | |
| BASE COUNT | 298 a 172 c 106 g 316 t 312 others | | |
| ORIGIN | | | |
| Query Match | 2.7%: | Score 39.2; | DB 121; Length 1204; |
| Best Local Similarity | 19.6%: | Pred. No. 5.7; | |
| Matches | 65; Conservative 134; Mismatches 131; Indels 2; Gaps 1; | | |
| Oy | 1135 | gatattcaatgtctgcataacccaacacttctaaccagtattacaataatagccttta | 1194 |
| Db | 1181 | KATAKAMKKAKAKKAAAAAATAAAAAAAAAAATAAATAAATAAATAAATAAADA | 1122 |
| Oy | 1195 | cagcagaatag--taggatacaagctactgttaaacatgcaaatggaggaaagtgttgga | 1252 |
| Db | 1121 | AADAAAARAKAKAKAAAAAADAAAMKKRAAAAAAAMAAAKMATTKDDDKDAKA | 1062 |
| Oy | 1253 | atggaatgactgcgcgttttatgtynaacaacggacatggtatatcatgctactcgaa | 1312 |
| Db | 1061 | DARKKKRKAADAKADADKDCKKAKADADMDADDAKKKKAKKKDKKDDATDADMAA | 1002 |
| Oy | 1313 | gcttttcgcgaattacggttcatactcgcgaacaagctactctagattatcataatgtcc | 1372 |
| Db | 1001 | KAAKKKDKKAAKKKKDKKRAKAKAKAKKKDKDKDKKKKKKKDKKAAKAAKAAADA | 942 |
| Oy | 1373 | aaaattcttctcaagtaaggaatggaggttcagttccatccagcgagtgatagtg | 1432 |
| Db | 941 | AAAAAAKKKKDKKDAKAAAKKKKDAKAKKKKDDWAAADADDAKAAKAAKAAKKK | 882 |
| Oy | 1433 | gacaacatcgatttatatgataatgtgatt | 1464 |

```

Db      881 KKDADKKAARDAAKKKKDKDRAAKAADKXNAARK 850
RESULT 15
LOCUS   A2138670
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
QY
Db
QY
Db
QY

```

Wed Dec 20 15:33:30 2000

us-09-339-159-1.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 03:17:03 ; Search time 4256.96 Seconds
(Without alignments)
1415.860 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470
Perfect score: 1380
Sequence: 1 aatgcaatcgcgatttta.....ttgataatgtgattgtgaa 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_em:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_em:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_v1:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pl4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pl5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_v11:*
73: gb_v12:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pl6:*
92: gb_pl7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------|---------------------|
| 1 | 458.6 | 33.2 | 1551 | 1 AB007123 | AB007123 Bacillus |
| 2 | 242 | 17.5 | 1530 | 2 D86329 | D86329 Vibrio sp. |
| 3 | 134 | 9.7 | 4567 | 1 AF163837 | AF163837 Calidpact |
| 4 | 131.8 | 9.6 | 2155 | 2 CDCMANABD | M36063 C.saccharol |
| 5 | 131.8 | 9.6 | 4977 | 2 CDCMANA | I01257 Caldoceillum |
| 6 | 125 | 9.1 | 5284 | 2 CSU16308 | U16308 Caldoceillum |
| 7 | 109.4 | 7.9 | 1461 | 5 E02075 | E02075 Genomic DNA |
| 8 | 109.4 | 7.9 | 1939 | 2 BACMANN | M31797 Bacillus sp |
| 9 | 94.4 | 6.8 | 25970 | 2 SC2H4 | AL031514 Streptomy |
| 10 | 89.2 | 6.5 | 1881 | 74 STMBMANASE | M92297 Streptomyce |
| 11 | 81.6 | 5.9 | 837 | 74 TFG6227 | AJ006227 Thermomon |
| 12 | 52.4 | 3.8 | 14922 | 2 SCF73 | AL121746 Streptomy |


```

Db      878  TGCGTATCTGATGATGACAAATGATTGGCGTGTAACTCCCTACCTCGTTCGGTATA 937
QY      842  caatagtgaatggttccatattggtttaagaagaaccttcgagattaaagcaccgtttttacag 901
Db      938  CCTAGTGAATGACAGTAAAGCACTTCTGTATTATCGGCAATTTTGGAG 997
QY      902  gtgagagatctgttgaggaactctccgcaactc----- 937
Db      998  GTGTTACGCCACCTCAAGCCCTACTTCTTACCTTACCTACCTCAACTCAACTCTCTTA 1057
QY      938  -----ttatgatttg 949
Db      1058  CTCCTAGGCCAAGTCGAGCCCGAGTCGAGTATAACGGGAGAGATCTATATGATTTCG 1117
QY      950  aagtgatgataaggaatgagctggaagtagcttgagcggaggtccttggtctgtgacag 1009
Db      1118  AAACAGGAATCAAGGCTGTCGGGAACAATATTTTCGGAGAGCCCTATGGGTCAACCAATG 1177
QY      1010  agtgcctcttaagaagaacatctttaaagcggatattcaattgctgcaaatcac 1069
Db      1178  AATGGAAGAACGGGAGCGCAACTCTCAAGCCGATGTCTCTTACCAATCCAAATCCA 1237
QY      1070  aacattacttaccatggttatacaaatcgtcttaccagcagaatagtagatacaagcta 1129
Db      1238  CGCATAGCTATATATATACCTCTAATCAAAATCTGTCTGAAAAAGCAAGTCAAAACCA 1297
QY      1130  ctgttaaacatgcaaatggaagtggttgtaatggaatgagctgcgtcttatttga 1189
Db      1298  CGGTTAAGCATGCGAACTGGGCAATATCGCAACGGGATTTATGCAAACTATACGTAA 1357
QY      1190  aaacgaagaactgttatacagtgtagctctggaagcttggcggatgaaggttaccctg 1249
Db      1358  AGACCGGGTCCGGGTGACATGTGTACGATTCGAGAGAAATGTGATTCATCAACGACG 1417
QY      1250  gaacacgcctatctcagatttaccatggtccaaatcttctcaagtagaaggaaattg 1309
Db      1418  GTACCAATTTTGAACATATCCCTCAGCGGCATTTGCAATTTGCTCCTCACTCAAGAAATTC 1477
QY      1310  gaattcagttccaatcagcagtgatagtagtggacaacaacatcgattatattgaaatg 1369
Db      1478  GGGTAAATATCCCGCTCCTCAACAGGTAGTGCCCAATACGATATTTATGTATGATGTG 1537
QY      1370  tgaattgtaga 1380
Db      1538  TTAGTCTGCA 1548

RESULT 2
LOCUS   D86329 1530 bp DNA BCT 19-MAR-1998
DEFINITION Vibrio sp. gene for beta-1,4-mannanase, complete cds.
ACCESSION D86329
VERSION D86329.1 GI:2978250
KEYWORDS beta-1,4-mannanase; manA.
SOURCE Vibrio sp. (strain:MA-138) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 1530)
AUTHORS Tamaru, Y.
TITLE Direct Submision
JOURNAL Submitted (01-JUL-1996) to the DDBJ/EMBL/Genbank databases. Yutaka
Tamaru, Mie University, Faculty of Bioresources, Utilization of
Marine Products, 1515 Kamihama, Tsu, Mie 514, Japan
(E-mail: s9tamay@maric.bio.mie-u.ac.jp, Tel: 0592-31-9561,
Fax: 0592-31-9557)
2 (sites)
REFERENCE Tamaru, Y., Arai, T., Morishita, T., Kimura, T., Sakke, K. and
AUTHORS Ohmura, K.
TITLE Cloning, DNA sequencing, and expression of the beta-1,4-mannanase
JOURNAL gene from a marine bacterium, Vibrio sp. strain MA-138
FEATURES
location/Qualifiers
1..1530

```

```

/organism="Vibrio sp."
/strain="MA-138"
/db_xref="taxon:678"
32..37
-35_signal
-10_signal
RBS
136..141
149..1336
/gene="mana"
149..1336
/gene="mana"
/codon_start=1
/translation="beta-1,4-mannanase"
/protein_id="BAA25188.1"
/db_xref="GI:2978251"
/translation="MKETRAIFSLLEFIWASCAHAGFYVSNGLVEANGSAPRKINGIN
HAHTWTDLKISVALSGIAATGANDYVAVLSNGYRWKNDVSDVTNINLAKANLITAI
LEHVDITGGESSAASLSDSADYVTELEKNEIGDEYVILNGNPFENNDAVAVY
NDHVSALORLSAGINHTIMVDAPNMGDMKPLMNAQFVNSDKLNTIFSVHME
VYSYNSVNDYISSFTNNGLVIGFPASTHKGADYDESIMERSSETLSIGYIGWES
GNDTTSDDIDIVNMNDNSYSTWGNVLIENGKIKSTSLATVFTGNDCCNDSSCEY
PICSSAAYDPDDDGWGMENNOSCIVODSSDITAPNGYPCQSSSDPDG6GWEMENNAS
MCC"
BASE COUNT 469 a 211 c 333 g 517 t
ORIGIN
Query Match 17.5%; Score 242; DB 2; Length 1530;
Best Local Similarity 57.2%; Pred. No. 4,9e-48;
Matches 525; Conservative 0; Mismatches 375; Indels 18; Gaps 4;
QY 3 tgcgaattccggaatttattgaagcgggtaccactctataagaatggcaatggcaaccatt 62
Db 202 TGCTCATGCGGGTCTTATATATCAATGCTGCTTATATGAAGCGAATGTGTGCTT 261
QY 63 tgaatgagaagattaaacattggcgcagcatggtataaagacaggaactactgcaat 122
Db 262 TAAATTTAGAGGATTAATATATGACATCTTGTATACGACAAAGTTAAGCTGCTTT 321
QY 123 tgaaggattgcaataccggtgcttaatacgcgtcggatgtgttactatgtagggagaca 162
Db 322 AAGTGTATGCTGCCACAGAGCAATATGTAAGGGGTGTACTAAGTAAAGTATGCT 381
QY 183 atggacaagaatgacatccatcagtaagaacctatctctttagcggaaagataatc 212
Db 382 GTGGACAAAATAAGACCTGTATGATGTGACTAATTAATTAATTAACAAACCAACAA 441
QY 243 ttggttcgtctctgaagttcatgattgtagtcggttattg-----atccattcg 293
Db 442 CTTTATAGCTAATTTAGAACTTATGACACTTACTGCGTATGAGAGAAAGTATGCTGTC 501
QY 294 ttgcctcaatcgtcgtgttattatttgaattgaatgagaagtgctttaaattgaaagga 353
Db 502 TAGTTTAGATTCGCGCGCTATACGTAGTGAATGGAAGAAGAGCTAATCGACAGA 561
QY 354 agataccgtcatattaatatttgcgaatgaatgtgttg---ttcgtgggaaggagatgc 410
Db 562 AGATTATGTCATCATTAACCTAGCAATGAACCTTTTGGTAATTAATTAATATCGGTAGC 621
QY 411 ttggcctgacggttataaagaacatcccgagattcgcgaacgctgtaaacctac 470
Db 622 TTGGGTTAATGACCATGTCTGCTTATCAACGATTAAAGAGCGAGTATCATTCACAC 681
QY 471 ctgatatgtagatgctgcgggtggtgggacaa---tttccacaatcgtatcatgattgg 527
Db 682 GATTATGTTGATGCTGCCAATGAGGGGCAAGACTGGAAGGGCTTATGCTGATATATGC 741
QY 528 aagaagaagtttttaagctcgaacctcaagaataacaaatgcttgcattcatatgtatga 587
Db 742 TCAATTCGTTTAAATTCGACCTTAATGAATGAATGATTTTATGCTTATATGATGA 801
QY 588 atatgcggtggttaatgacatcgcaagttctgactaatattgaccgagtttcaatcaga 647

```

| Db | 802 | GGTTATATAGTACAAATTC--AGTCATGACNACATCTCATATTACAAACAATG | 858 |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-----------------|
| QY | 648 | ccctgcattagtcattggtgaaatttggaaacccgcatcacataatgtatgcctgataagc | 707 |
| Db | 859 | CTTAGCGTGGTTATTCGGGTGAGTGTGCTTCAACTATAAAGGGCCGAGTGTGATGAAAG | 918 |
| QY | 708 | aacgattatgagctatctcgaaacaagagagttgggtgtgttggcgtggtcatgaaag | 767 |
| Db | 919 | TTCATTAATGAAACGTTTCAGAAACGTTAGTTAGTTATTCGGTGTCTGTGTCAGG | 978 |
| QY | 768 | gaacggcccaagaatggagatttagaccttcgaatgtatggcttggcgaataacctac | 827 |
| Db | 979 | AAATGATATACACATCCGATTCGATTGTGATTTGTTATATATGTTGGAGCAATAATTCGATAG | 1038 |
| QY | 828 | agcttggggaatacacatgatagtatgcatatggttttaagaagaactcgagattag | 887 |
| Db | 1039 | CACTTGGGTATGTTTATTATTAATGATCAAAATGTATTAATTCACATCACTGACGTTAGC | 1098 |
| QY | 888 | cacgcgtttttacaggtg 905 | |
| Db | 1099 | TACGGTGTACTTGTG 1116 | |
| RESULT | 3 | | |
| LOCUS | AF163837 | 4567 bp DNA | BCT 08-FEB-2000 |
| DEFINITION | Caldibacillus cellulovorans multidomain beta-1,4-mannanase precursor (mana) gene, complete cds, and unknown genes. | | |
| ACCESSION | AF163837 | | |
| VERSION | AF163837.1 | GI:6651325 | |
| KEYWORDS | | | |
| SOURCE | Caldibacillus cellulovorans. | | |
| ORGANISM | Caldibacillus cellulovorans | | |
| REFERENCE | Bacteria; Firmicutes; Bacilli/Clostridium group; Bacterium/Staphylococcus group; Caldibacillus. | | |
| AUTHORS | 1 (bases 1 to 4567) | | |
| TITLE | Sunna, A., Gibbs, M.D., Chin, C. W.J., Nelson, P.J. and Bergquist, P.L. | | |
| FEATURES | A gene encoding a novel multidomain beta-1,4-mannanase from Caldibacillus cellulovorans and action of the recombinant enzyme on kraft pulp | | |
| JOURNAL | Appl. Environ. Microbiol. 66 (2), 664-670 (2000) | | |
| MEDLINE | 20120520 | | |
| REFERENCE | 2 (bases 1 to 4567) | | |
| AUTHORS | Sunna, A., Gibbs, M.D. and Bergquist, P.L. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (29-JUN-1999) Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia | | |
| FEATURES | location/Qualifiers | | |
| source | 1. 4567 | | |
| CDS | /organism="Caldibacillus cellulovorans" | | |
| | /db_xref="taxon:74586" | | |
| | <1..513 | | |
| | /note="orf1" | | |
| | /codon_start=1 | | |
| | /translation="unknown" | | |
| | /product="unknown" | | |
| | /protein_id="AA022273.1" | | |
| | /db_xref="GI:6651326" | | |
| | /translation="TPTPTPSAAPPTPSAGSLVYGRADTNAGDNLPHFRI | | |
| | VNGTSVPSLELTIWYVDGKDPVENCMDWAQVCSNRSFVLSGRGADY | | |
| | IELTTSVAGSLAAGASSGDIOVINRNDWTINNEANDYSIDPTKSFADNRVTLIR | | |
| | NGDLINQVDP" | | |
| | 1..60 | | |
| misc_feature | /note="Region: proline-threonine linker" | | |
| | /evidence=not_experimental | | |
| | 61..510 | | |
| misc_feature | /note="Region: cellulose-binding domain type IIIB" | | |
| | /evidence=not_experimental | | |
| | 620..3412 | | |
| gene | /gene="mana" | | |
| | 620..718 | | |
| sig_peptide | /gene="mana" | | |
| | /evidence=not_experimental | | |

| Query Match | Best Local Similarity | 9.7% | Score 134 | DB 1 | Length 4567 | | |
|-------------|------------------------------------------------------------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|------------------------------|-------------------------------------------|-----------|
| Matches 408 | Conservative | 0 | Mismatches 360 | Indels 18 | Gaps 4 | | |
| 68 | tgagaggattaacaccatggtgacgcacgtgtatataaagaccagcaactactctgaattgaag | 127 | 620.3412 | /gene="mana" | /codon_start=1 | | |
| 2070 | tggttggtattaaacaccacccgcatcggtgattccgacacggctacgttgcgtggtgcaag | 2129 | /product="multidomain beta-1,4-mannanase precursor" | /protein_id="AAF22274.1" | /db_xref="GI:6651327" | | |
| | | | /translation="MNRRLRLRLSGMLAMVLIAMLAIVKPPPEAEHAGHGVPAATPTT IACVADKGVHNGGDDLNINIPACIDALAIISGNTQFVWMPFGNLISNAGGRHEIIIPDK LCGLPASPDMNQARIDMTWMTIRLQPGATITVRYVAMAPHPGTWLYTRDGMPTQPL KMSDLPEPVSQVTPNIPINSISGPDGAESMQVQPLNKGRIIIMIQRSDSPAEVYN CSDFVGGSPFAVEFGDPRREGTMTTPPSQVTPPTPTPTPTPTPTPTPTPTPTPTPT TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT STOAEQPCDDMAQIGCSNINRAQFYSLSQPPSAGSYIELSTGSGISAGNTGIGTOR IHFNNMANYNETDMSTINGAQTIWSPGSTRILINGLVWETEGGSSPTPTPTPTPT PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT GAMVRIYLSNGCRMTKIPASEVADIIISQARTLQRAVVLVEHDTTGYGEDAASMT TAVVWIELKNVLVAQENFVIVNIGNPPYGNNNQVNTDTRNAVQSLNAGINNTIM VDANMGQDNSTFTRDNAPITFNADPQNRNVFSTHMTGVYDTAEVQSYIESFNRIL PLVGEFGHMSDGPDPDQALVYOAKYNGIKFEGMSGGNGGVETDTPNANP TAKGTERTNAIGRTSPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT KPHRIYNRGTSVYPSLELIRITWYTDGDPQVFNCDMAQVGSNIRGFSVKLSTGR TGAIYILEITITSGAGSLAPASSGDIQVIRINKNDWYNNEADYSYDPKTSFADWN RVLYIRNGQLWVGVEP" | 719.3409 | /gene="mana" | /product="multidomain beta-1,4-mannanase" | 1349.1480 |
| | | | /gene="mana" | /note="Region: proline-threonine linker" | /evidence="not_experimental" | | |
| | | | /gene="mana" | /note="Region: proline-threonine linker" | 1481.1927 | | |
| | | | /gene="mana" | /evidence="not_experimental" | 2033.2884 | | |
| | | | /note="Region: cellulose-binding domain type IIb" | /evidence="not_experimental" | 1928.2032 | | |
| | | | /gene="mana" | /note="Region: proline-threonine linker" | 2033.2884 | | |
| | | | /evidence="not_experimental" | 2859.2959 | | | |
| | | | /evidence="experimental" | 2885.2959 | | | |
| | | | /gene="mana" | /note="Region: proline-threonine linker" | 2960.3409 | | |
| | | | /note="Region: cellulose-binding domain type IIb" | /evidence="not_experimental" | 3894.4400 | | |
| | | | /note="orf3" | /codon_start=1 | /transl_table=11 | | |
| | | | /product="unknown" | /protein_id="AAF22275.1" | /db_xref="GI:6651328" | | |
| | | | /translation="MGASGVFSRTAONGYGLKARCQRRVCVGNHLGHLFVDDN RFGAGAKRRLGRFSRQNRNLGEAONFESGVTFVQDDDKRIRRENIGOLFVYD GPKRLIEIAQROBARAKARAKQASDQPLQKNVILHFIIDQHMVGYVERMDPFRHKE RSILTRYV" | 1276 | 1439 | | |
| | | | 812 | t | | | |
| | | | | | | | |

| DB | 1655 | TGTCAGGATGCAAAACGATCAAGATAGCATTTTACTGTGGTCGTCGTGTGAATTT | 1714 |
|----|------|---------------------------------------------------------------|------|
| QY | 773 | gccccgaatgggaagtatttagaccttcgaatgatgtggcctggaaataaccttacagctt | 832 |
| Db | 1715 | CGAGCTATGTGGGATTTTGACATGCTAACCACTGGGACCCCAATATCAACTCAT | 1774 |
| QY | 833 | ggggaaa | 839 |
| Db | 1775 | GGGGACA | 1781 |

| | Query Match | Best Local Similarity | Matches | Conservative | Score | DB 2 | Length | Indels | Gaps |
|-----------------------|---------------------------------------------------------------------|---------------------------------------------------------|----------------------------|-----------------|------------|------|--------|--------|------|
| misc_feature | /note="encodes proline-threonine linker number 2" | 2938. .3419 | | | | | | | |
| misc_feature | /gene="celc" | /note="encodes cellulose-binding domain (CBD) number 2" | 3420. .3524 | | | | | | |
| misc_feature | /gene="celc" | /note="encodes partial mannanase domain" | 3420. .3524 | | | | | | |
| misc_feature | /gene="celc" | /note="encodes proline-threonine linker number 3" | 3574 | | | | | | |
| misc_feature | /gene="celc" | /note="location of the first -1 frameshift" | 3575. .3625 | | | | | | |
| misc_feature | /gene="celc" | /note="ORF #2 from frameshift #1 to frameshift #2" | 3627 | | | | | | |
| misc_feature | /gene="celc" | /note="location of the second -1 frameshift" | 3628. .4374 | | | | | | |
| misc_feature | /gene="celc" | /note="ORF #3 from frameshift #2 to celc stop codon" | 1829 a 881 c 1238 g 1336 t | | | | | | |
| BASE COUNT | 1829 a | 881 c | 1238 g | 1336 t | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match | 9.1% | Score 125; | DB 2; | Length 5284; | | | | | |
| Best Local Similarity | 51.1% | Pred. No. 7.6e-20; | | | | | | | |
| Matches | 405; | Conservative | 0; | Mismatches 370; | Indels 18; | Gaps | | | |
| QY | 121 attgaaggatgccaataaccggtgctaatacgcgtcgatgtgtatctgaatgggga | 180 | | | | | | | |
| DB | 3610 ATTGCGTGAATCAGCAGCATGGGGTATGAACTCTGAGAGGGTACTGATGATGTTAC | 3669 | | | | | | | |
| QY | 131 caatgacaaaagaatgacatcatcagtaagaaccattatctcttgcggaagatat | 240 | | | | | | | |
| DB | 3670 CGATGACGAGAAATACCAAGCAAGTAAAGTGGCAATATATATATCTTTCAGGAAGTCTT | 3729 | | | | | | | |
| QY | 241 catgtgttcgtcttcttgaagttcatgagttacccgttatgattccatg | 292 | | | | | | | |
| DB | 3730 GGTTCAAAGCATATATATTAGAACTAACACACAAAGATATGGAAGATGGGCA | 3783 | | | | | | | |
| QY | 293 ---cttcgcctcaatcgctgctgtgtatattgattgattgaatgagaagtccttaattgga | 348 | | | | | | | |
| DB | 3790 GCGTGTCAATGATGATACAGCAAGTGAATATTTGGAAGGATTAAGAGCGTATTAACGCT | 3849 | | | | | | | |
| QY | 349 aaggaagataccgctcatatataatttcggaatgaaatggttgcgttgcgtggaaaggat | 406 | | | | | | | |
| DB | 3850 AACGAAAGATTTTGTATATTAATTAACATTTGGCAATGGCCGTATGGGACAAATTAACATTA | 3909 | | | | | | | |
| QY | 409 gcttgagctgcgggtatataaagaagcaatcccgagttcgtaagcgcgcttaagcat | 468 | | | | | | | |
| DB | 3910 AACTGGATTATACACGAAAGACGCTATTAAGCACTTAGAGATGCGAGATTCAAGCAC | 3969 | | | | | | | |
| QY | 469 aacctgattgattagatgctgcgggtgaggagac---aattccacaatcgattcatgattat | 525 | | | | | | | |
| DB | 3970 ACGAATATGAGGATGAGCGCCCAACTGCGGCTCAGATTTGGTAAATACATAGAGATTAAT | 4025 | | | | | | | |
| QY | 526 ggaagaagaattttaatgctgagccctcaacgaataatacaatgatttcgattcataatg | 563 | | | | | | | |
| DB | 4030 GCCCGACAAATATGGAAGCAAGATCCCTGGCGCATTTGTATTTGCATTCATATATAT | 4089 | | | | | | | |
| QY | 586 gaattgagtgatgaaatgacatcgaaagtgtgtaattatgaccgggttccttaataa | 645 | | | | | | | |
| DB | 4090 C---GTGATATCAATACAGCAAGCAAGGTGCAAGATGATCAATCAATTTGTGATAG | 4146 | | | | | | | |
| QY | 646 gacctgcatatgattcattgctggaatttggacaacgcttacaatgagtgagtcgattgaa | 705 | | | | | | | |
| DB | 4147 GGGTACCATTTGTTATTTGGGGAATTTGGACATACGACACAGATGGTACCCCTGATGA | 4206 | | | | | | | |
| QY | 706 gcaacgattatgagctattctgaacaagaaggagttgggtgtgtgctggtcatgga | 765 | | | | | | | |
| DB | 4207 GAACGATATGACGATATGACAAACAGTACAGATATGATATTTAGTTGGTGGCGGTGT | 4266 | | | | | | | |

| | | | |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|---------------------|
| OY | 766 | gggaacgccccgaatgagatgttatagccttcgaaagtattggcctggaataaccct | 825 |
| Db | 4267 | GGAATTTGCAAGCTAATTGGTGGATTGTTCACATGATAAAACAAGTGGACCCCAATTAATCCA | 4326 |
| OY | 826 | acaagtttgagggaataacataagtagtatggtccatatggtttaagaagaacctcgagatta | 885 |
| Db | 4327 | ACTCATCGGGGGCACAATGATATAAACTAATGCAATTGATACAAAGAATGAATATAATA | 4386 |
| OY | 886 | agcacccgtttta 898 | |
| Db | 4387 | TTCAAAAGTGTTAA 4399 | |
| RESULT | 7 | | |
| LOCUS | E02075 | 1461 bp | DNA PAT 29-SEP-1997 |
| DEFINITION | Genomic DNA encoding beta-mannanase of Alkalophilic Bacillus sp. AM-001. | | |
| ACCESSION | E02075 | | |
| VERSION | E02075.1 | GI:2170317 | |
| KEYWORDS | JP 1989228477-A/1. | | |
| SOURCE | unidentified. | | |
| ORGANISM | unclassified. | | |
| REFERENCE | 1 (bases 1 to 1461) | | |
| AUTHORS | AKINO,T. and HORIKOSHI,K. | | |
| TITLE | BETA-MANNASE GENE DNA, RECOMBINANT PLASMID CONTAINING SAID DNA AND TRANSPORTANT | | |
| JOURNAL | Patent: JP 1989228477-A 1 12-SEP-1989; | | |
| COMMENT | RES DEV CORP OF JAPAN, AKINO TOSHIRO, HORIKOSHI KOKI OS Alkalophilic Bacillus sp. PN JP 1989228477-A/1 PD 12-SEP-1989 PF 09-MAR-1988 JP 1988053774 PI AKINO TOSHIRO, HORIKOSHI KOKI PC C12N15/00,C12N1/20//C12N9/24,(C12N15/00,C12R1:07),(C12N1/20, PC C12R1:19); CC strandedness: Double; CC topology: Linear; CC hypothetical: NO; CC anti-sense: No; CC "source: strain-AM-001; FH Location/Qualifiers FT CDS 1..1461 /product='beta-mannase of Alkalophilic FT FT FT Bacillus sp.'. Location/Qualifiers FEATURES source 1..1461 Location/Qualifiers /organism="unidentified" /db_xref="taxon:32644" | | |
| BASE COUNT | 475 a | 257 c | 340 g 389 t |
| ORIGIN | | | |
| Query Match | 7.9%; | Score 109.4; | DB 5; Length 1461; |
| Best Local Similarity | 55.0%; | Pred. No. 4.4e-16; | |
| Matches 259; Conservative | 0; | Mismatches 206; | Indels 6; Gaps 2 |
| OY | 910 | tctgatgaggggaactctccgacaactctttagatcttgaagtgatgatcaagaatgag 969 | |
| Db | 991 | TATGGCAATCAATCAATCACGCACATGCTCTCTAATGATTGAAAAACAATAGCTATGCTGG 1050 | |
| OY | 970 | actggaagtagct---gagcggaggtgcccttggtggcggtgacagagtggtctttaaagga 1026 | |
| Db | 1051 | TCCGGGTGTGAATTTACGGACGCGAGAACCATGACATTCGAATTAAGTAGTGGCAATGTR 1110 | |
| OY | 1027 | agtcactctttaaagcggatattcaattgctgcaccaatcacacattacttacatgct 1086 | |
| Db | 1111 | ACTCAATCGTTGAAGACAGATGTCGTTCTGCGGCAATTAATAGTACCAT---TTGCAAAAA 1167 | |
| OY | 1087 | attcaaatagctctttacagcagaatagtaggataataagaagctctgtttaaactgcaat 1146 | |

| Db | 1168 | ACAGGATCGCAAAATCTAGTTCATTCAAAACACGACGAAATTAAGTAGGACATCTTCG | 1227 |
|-------------|------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|--------------------|
| Oy | 1147 | tggggaaagtgttgttaataaggaatgactgcgcgctcttatttggaaanacggacatggttat | 1206 |
| Db | 1228 | TGGGGAATGTAGGAAAGTGGCATGACACACAGAGTGTTCGTAACAAACAGGAGTCTTGG | 1287 |
| Oy | 1207 | acatgtactctcggaaagttcgtgcgcgaattcaaggttcacatcgtgaaacaagctatctcta | 1266 |
| Db | 1288 | AGATGGAATGCAAGCGAATTTTGTCAGTTTGCAAGGCAACACAGACGCACTACTACTT | 1347 |
| Oy | 1267 | gattatcaaatgtccaaatcttctcctaagtaagggaatgtgaggttcagttccaatca | 1326 |
| Db | 1348 | GATTGACGAAAGTAGAATTAATCTGCATGATGTTCGACGAGATAGGTGAGATTAACGA | 1407 |
| Oy | 1327 | gcgagtgtagtagtgcgcacacatcgattatattgtgtaattgtagtga | 1377 |
| Db | 1408 | CCAGCAATACGACACGGGAGACGGCGATTTACTTAGATCATGTACGCGTA | 1458 |
| RESULT | 8 | | |
| LOCUS | BACMAN | 1939 bp | DNA |
| DEFINITION | BACMAN | | BCI |
| ACCESSION | M31797.1 | GI:143166 | |
| KEYWORDS | 1,4-beta-D-mannan mannanohydrolase; mannan endo-1,4-beta-mannosidase; mannanase. | | |
| SOURCE | Bacillus sp. (strain AM-001) DNA, clone pMAH5. | | |
| ORGANISM | Bacillus sp. | | |
| REFERENCE | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus. | | |
| AUTHORS | Akino,T., Kato,C. and Horikoshi,K. | | |
| TITLE | Two bacillus beta-mannanases having different COOH termini are produced in Escherichia coli carrying pMAH5 | | |
| JOURNAL | Appl. Environ. Microbiol. | 55, | 3178-3183 (1989) |
| FEATURES | | | |
| source | Location/Qualifiers | | |
| | 1..1939 | | |
| | /organism="Bacillus sp." | | |
| | /db_xref="taxon:1409" | | |
| sig_peptide | 220..297 | | |
| | /note="beta-mannanase signal peptide" | | |
| CDS | 220..1761 | | |
| | /note="beta-mannanase precursor (EC 3.2.1.78)" | | |
| | /codon_start=1 | | |
| | /transl_table=1 | | |
| | /protein_id="AA02586.1" | | |
| | /db_xref="GI:143167" | | |
| | /translation="MKYAKVAFYMAIFMFSVLPTISMSEANGAALSNPNANOITTY | | |
| | NVYSMLANLPMKRSRYVSGHFHGYSPSTLMIKOCARLEGKMPGILSCDYKNNQCFE | | |
| | LYVADQSLYSGNDELINFMNQGGLVTTISHNPENGFHGEYVKTILPTSPQNLNTH | | |
| | TTBERRKKMDKKAADGIDELONGGVLYLPELHMNGEMWMAEGENORFORAN | | |
| | YISWRKMYQVTFHERKLTNNLTIMYSPDYVDYHTSYYPGANYDIYALDSYHPDH | | |
| | LTDYIKRIALDKPFAREIGPESMAGSPDYSNTQAIKQKIPRYTYFLANDKMS | | |
| | HANGGADLENDWSVYVRNGELIDYGOSNPATLYLDFENNTLSWSCETFDGPMWISN | | |
| | SNNSQIDKADVYVGNNSYIKQKTVNRNLSFKNLEIKVSHSGNVSQSAVYV | | |
| | TGSMRWNAEGFCQFAGKRTALSIDLTKYNSLHDVREIGVEYKAPANSCKKIAL | | |
| | HYVTR" | | |
| mat_peptide | 298..1758 | | |
| | /note="beta-mannanase" | | |
| misc_signal | 1826..1840 | | |
| | /note="pot. transcription termination signal; putative" | | |
| misc_signal | 1861..1889 | | |
| | /note="pot. transcription termination signal; putative" | | |
| misc_signal | 1905..1916 | | |
| | /note="pot. transcription termination signal; putative" | | |
| BASE COUNT | 637 a 343 c 420 g 539 t | | |
| ORIGIN | 1 bp upstream of XbaI site. | | |
| Query Match | 7.9% | Score 109.4; | DB 2; Length 1939; |


```
Db 25173 ACGGCCAGCTCCCTGCGGCGGTACCTCGCTGCTGCTGAGAGGCGGCGGCGGCGC 25232
Qy 781 tggagattatgaaccccttgatgattggtcggaataaccccttgatggaat 840
Db 25233 GTCAGATCCTCAGACATGCTGACGGCTTCAGACCGAAGCTGACGAGCTGGGGCAGC 25292
Qy 841 acaatagtaagtgcataatgattagagaactcgagaattacgacgatttaca 900
Db 25293 CGCATCTTCTCAGCAGACGCGATCGCCGCGACGCTCAGAGCGCCACCGTGTACGCG 25352
Qy 901 ggtggaag 908
Db 25353 GGTGTGG 25360

RESULT 10
STMBMANASE 1881 bp DNA BCF 16-JUN-1999
LOCUS STREPTOMYCES LIVIDANS mananase (mana) gene, complete cds.
DEFINITION M82297
ACCESSION M82297.2 GI:4579683
KEYWORDS
SOURCE Streptomyces lividans.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 1881)
AUTHORS Arcand,N., Kluepfel,D., Paradis,F.W., Morosoli,R. and Shareck,F.
TITLE Beta-mannanase of Streptomyces lividans 66: cloning and DNA
sequence of the mana gene and characterization of the enzyme
JOURNAL Biochem. J. 290, 857-863 (1993)
MEDLINE 93207541
REFERENCE 2 (bases 1 to 1881)
AUTHORS Shareck,F.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1992) Centre de recherche en microbiologie
appliquee, Institut Armand-Frappier, 531 boul. des Prairies, Laval,
Quebec H7N 4K3, Canada
REFERENCE 3 (bases 1 to 1881)
AUTHORS Shareck,F.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1999) Centre de recherche en microbiologie
appliquee, Institut Armand-Frappier, 531 boul. des Prairies, Laval,
Quebec H7N 4K3, Canada
REMARK Sequence update by submitter
COMMENT On Apr 8, 1999 this sequence version replaced gi:153193.
FEATURES
source
1..1881
/location/Qualifiers
1..1881
/organism="Streptomyces lividans"
/strain="1326"
/db_xref="taxon:1916"
217..1316
/gene="mana"
217..221
/gene="mana"
/feature="putative"
226..1377
/gene="mana"
/EC_number="3.2.1.78"
/codon_start=1
/translation="11"
/product="mannanase"
/protein_id="AA26710.2"
/db_xref="GI:4579684"
/translation="MNRARSTLTITAGMAFVILGLFALAGPSRAEAAGIHYNS
GRVVEGSAFVKGVNHATYTPDRTGSIAIAKAGANTVRYVLSGGHWTTSASE
VSAIIGCKKANKYICVLEHDTGYGKGATSIIDAGDYVGVKSAMRQEDYVYVN
IGNEPFGNTYAAATDATKSAIGLRKAGIGHALMDAPMWDGDSGTASNAAYTA
SDPDRNTVESIHMTGVDTAEVADYINATVNGCLPTVEFGDQSHDDEDAIYA
TAOSLAVGLMGMSGNGGVEYIDMNVNGDPSNLSWGNRIIYSGNSGIAITSTATV
YGGGGSTGGTAENGYPYCVNGASDEPDGDGMWENSRSCVVGASADH"
226..333
sig_peptide
```

```
mat_peptide /gene="mana"
334..1374
/EC_number="3.2.1.78"
/function="hydrolase"
/product="mannanase"
BASE COUNT 309 a 648 c 647 g 277 t
ORIGIN
Query Match 6.58; Score 89.2; DB 74; Length 1881;
Best Local Similarity 47.68; Pred. No. 3.2e-11;
Matches 432; Conservative 0; Mismatches 458; Indels 18; Gaps 5;
Qy 13 ggaattatgaagcgggtaccacccctatagatgcgaatggaaccattgtaatga 72
Db 340 GGCATCCAGCTCAGACACGACGCTGCTCGAGGCAACGCGCGCTGTCTATGCGC 399
Qy 73 ggaattacacatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 132
Db 400 GGGGCAACACGCGCTACACCTGATACCGAC---GCACGGGCTCATGCGGACATC 456
Qy 133 gcaaataccggtgctaatacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 192
Db 457 GCGGCCAAGGAGGCAACACCGCTCGCTCGCTCAGCAAGCGCGCGCGCTGAGCAG 516
Qy 193 gatgaatccatgaagaaacattatctcttagcggagaataatcattggtgct 252
Db 517 ACGAGCGGCTCAGAGCTCGGCGCTCTCGCCAGTCAGCAAGGCCACMAAGTATCTGT 576
Qy 253 gttcttgaagtcataatgctaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 306
Db 577 GTCCTTAGAGTGCACAGACACACCGGCTACGCAAGAGCGCGGACCGACGCTGAGCAG 636
Qy 307 gctgttgattatgattgaatgaagatgct---ttaatggaaggaataacgctc 363
Db 637 GCCGCGACACTAGTGGTGTGCTGTAAGAGCGCCCTGGAAGGCGCCAGGAGTACTG 696
Qy 364 attattatattggaatgaatgattggttcgcgggaagagatgcttgctggcgaagg 423
Db 697 GTCGTAAACATCGGACAGACCCCTTCGGACACCAACTACGCGCGCTGACCGCGG 756
Qy 424 tataacaagaacatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 483
Db 757 ACGAAGTGGGCAATCGGACACTCGCGCGCGGCTCTCGGCAACGACCTCATGCTGC 816
Qy 484 gctcgggggtgggac---aattccaacatcgatcattgattggaagaaagatttt 540
Db 817 GCGGCCAACTGGGCGCAGAGACTGCTCGGCAAGATGCGGTCAACGCGGCGCTCGT 876
Qy 541 aatgtgacctcaagaataataatgcttctgattatgattgattgattgattgattg 600
Db 877 GCGTCCGACCCGACCGACACCGCTTCTCCATCCACATCGATCG---GGGTACGAC 933
Qy 601 aatgatcgcaagcttgctactaatattgacccgaggtctttaaagaacccctgcatg 660
Db 934 ACCGCGCGGAGTGGCGGAGTACCTTAACCGCTTCTCGGCAACGACCTCCATGCTC 993
Qy 661 attgtgaatttgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 994 GTCGGCGAGTGGCGGACACACAGTACGCGCAACCCCGACGAGACGCATCATGCGC 1053
Qy 721 tatctgaacaagaagagattggtggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 1054 ACGGCCAGCTCCCTGCGGCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1113
Qy 781 tggagattatgaaccccttgatgattggtcggaataaccccttgatggaat 840
Db 1114 GTCAGATCCTCAGACATGCTGACGGCTTCAGACCGAAGCTGACGAGCTGGGGCAGC 1173
Qy 841 acaatagtaagtgcataatgattagagaactcgagaattacgacgatttaca 900
Db 1174 CGCATCTTCTCAGCAGACGCGATCGCCGCGACGCTCAGAGCGCCACCGTGTACGCG 1233
```


JOURNAL
MEDLINE
REFERENCE
AUTHORS
JOURNAL
3 (bases 1 to 14922)
Unpublished
3 (bases 1 to 14922)
Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (04-OCT-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>)
CDS are numbered using the following system eg SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nh1.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid F73 Overlaps with cosmid F11 on the BseI-F genomic
restriction fragment.

FEATURES

source

1. 14922
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid F73"
1. 122
/note="nominal overlap with StfII"
complement(1..983)
/gene="mana"
complement(1..983)
/gene="mana"
complement(1..983)
/note="SCF73.01c, mana, probable secreted
beta-mannosidase, len: 327 aa; similar to many e.g.
MANA, STB1, P51529 mannan endo-1,4-beta-mannosidase
precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189
z-score: 1332.7 E(1): 0, 61.1% identity in 288 aa overlap.
Contains N-terminal signal sequence, and Pfam match to
entry PF00150 cellulase, Cellulase (glycosyl hydrolase
family 5). Also similar to SC2H4.16 (62.2% identity in 288
aa overlap)."
/codon_start=1
/transl_table=11
/product="putative secreted beta-mannosidase"
/protein_id="CAB57406.1"
/db_xref="GI:6013077"
/translation="MRPARDPRTLPRLNGIALIGLVVVGALCPGALAGSPRA

CDS

RBS
gene
CDS

RSAGVPSAGQAAGLHIGDRLLGNGDNFVARGVNHATHTYTGSTOSLADYKALCAN
SVRYVLDGHRNSENPGADVAAYVTGCKANRLICYLEVDHTDGAAGAAGLLDAD
YKIGLADYLAGGEDYIYVINGNPNRNTDPACHTPEYVAVKKRAAGQRIIMDAD
NMGQDVYARAKASVYDADPTGNLISITHEI SVFDIAQETIDYINAAVDELPLILI
GEGFGADYGDPEDTMMATAEQRLGYLAWMSWGSNTDPVIDLADFDPS"
complement(989..992)
complement(1129..2724)
/gene="SCF73.02c"
complement(1129..2724)
/note="SCF73.02c, possible membrane-bound oxidoreductase,
len: 531 aa; some similarity to eukaryotic
molybdenum-containing oxidoreductases e.g. SNOX RAT 007116
sulfit oxidase precursor (488 aa), fasta scores: opt: 262
z-score: 281.9 E(1): 2.5e-08, 30.4% identity in 319 aa
overlap, and N1AL_ORYSA nitrate reductase 1 (916 aa),
fasta scores: opt: 247 z-score: 262.1 E(1): 3.1e-07, 28.9%
identity in 336 aa overlap. Contains several membrane
spanning hydrophobic domains near N-terminus, and Pfam
match to entry PF00174 oxidored_molpb, Oxidoreductase
molybdopterin binding domain"

gene
CDS

/codon_start=1
/transl_table=11
/product="putative membrane-bound oxidoreductase"
/protein_id="CAB57407.1"
/db_xref="GI:6013078"
/translation="MSDHKTRNRSARFTWTLRGALSGLLAGCAALAVETLNRV
PNSGPVAVAGAIIDPTPAVDMALIRFGTIDKLVLOGLIAYLLEFLAAGNATV
HRGAGAGLVGIVGAAALGRPDSTGTVDPSPVIGIAVALYLGLAGLLEPFL
EABERGECDMRREYIATAAASAGTGVGRALSGAGREAVASRENTILPPDS
RNPAYPROAKAREVGSSTPTSDPRTDPAVYKVAATAMRLRIHGEVAREBTI
TDDILRLRLIDITLTCVSNBVGSPYGNARNICVRLADLACGVPRPAGGACD
QVARSVDGMTIGSPVEDWDRDLAAGNGEPLPEDHGFVRRVVGGLGIVSAC
KMIIDEIETTFDSYDAVYKRWMPAPVKTQRIDTPRPRPAGTVAGVAMQ
HWIIDEIVRVDDGEWEAVLAAEESRDTWRMSYAMRATKGGHLLTVALLDRAISVY
TKRRRTVVDGASGRHSVYTVD"
complement(2777..4300)
/gene="SCF73.03c"
complement(2777..4300)
/note="SCF73.03c, unknown, len: 507 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCF73.03c"
/protein_id="CAB57408.1"
/db_xref="GI:6013079"
/translation="MTAFILVSGMFTGTHIMQDTAARLTARGAEVHTVALTGIDAPRA
AAAPGVDETHADYLAIVDSVGAAGDRITVVGHDYGIHPAVGAADRAREVDRIH
IDSGPPOCVPRALAVPPOSLERLAAGAGAGCVYPPRAHEMPPMGSTAGVPDADLD
RLTALAAPQPLSTLQPLRLTGAVDPVPTGVLTGNGSTIELVQMLVRLGDPALPL
TDPVRSFEELPTGHPMPLSLPAELTDVYLRAAGAGRRLEPDDTADPGHLPFLMDY
PDVPRERHGNLDLYLPDAEPPAVVLVVGSPVADARTPPDMGGLGYARCAVGDG
AVGALLDHLHDLDGFERAAADVAAVEVRAVDPRDGRVRLVLFSSGGGLAAADMD
APPAMRLCIAATVPIALPLPMNGLSERPRPRAVANGALSVLTTRVCREMPLAT
VEFLAAKDCADVVDVDPGHGHGFEITIDTDSRTAVRAATVLDHAGGAR"
complement(4311..4315)
4382..4387
4392..5414
/gene="SCF73.04"
4392..5414
/gene="SCF73.04"
/note="SCF73.04, probable transcriptional regulator, len:
340 aa; similar in N-terminus to members of the Merr
family e.g. NOXA-BRN nodulation protein NOXA (237 aa),
fasta scores: opt: 211 z-score: 233.9 E(1): 1.2e-05, 48.5%
identity in 68 aa overlap. Also similar to SCD17.06c
(34.8% identity in 322 aa overlap). Contains Pfam match to
entry PF00376 merr, Bacterial regulatory proteins, merr
family"
/codon_start=1
/transl_table=11
/product="putative Merr-family transcriptional regulator"
/protein_id="CAB57409.1"

RBS
RBS
gene
CDS

```

/db_xref="GI:6013080"
/translation="METSNDGICVIGIELAAGAWYKTYREYSDRLLEASRSAGSH
RTGEPALDRILRLNSRLALGCVIPYVRILDRHRAHGTAGEGGLLEAAYGRLEAY
GSELRALRWREARLRLVQEPPEQADRLRLIGAVSPQSTAVLEFWGWM.PARRPA
RSNAGFLDAVAVPDPDEPAPAVQYLAFAALNALIACRGTVQPODEARRAAGARPA
LYTAGFLDAVAVPDPDEPAPAVQYLAFAALNALIACRGTVQPODEARRAAGARPA
DPLDGYWELVAEVTAPGARPEPTPGASHMDMLAALGTGAVPQGSPPQSKRRY
SPGLD"
gene
complement(5357..5623)
/ gene="SCF73.05c"
/ complement(5357..5623)
/ gene="SCF73.05c"
/ note="SCF73.05c, unknown, len: 88 aa"
/ transl_table=1
/ product="hypotheetical protein SCF73.05c"
/ protein_id="CAB57410.1"
/db_xref="GI:6013081"
/translation="MDEVEMRGVYGAHDNPGPAGRYAHLVGGPLDLLDVTD
LTERARAGVTLATETGRGPGSHAVTPRPGDTRFRDMGDP"
gene
complement(5696..6697)
/ gene="SCF73.06c"
/ complement(5696..6697)
/ gene="SCF73.06c"
/ note="SCF73.06c, unknown, len: 333 aa"
/ codon_start=1
/ transl_table=1
/ product="hypotheetical protein SCF73.06c"
/ protein_id="CAB57411.1"
/db_xref="GI:6013082"
/translation="MLNDPEVTVDRGRYRDTVPWEPAWTEALGMIENRLAAGLR
PGRWVRRLRPMVILRLCYAGDVAWKEKSPAGAEALITSLARWGHVLELA
VDAKGMVILPDGGLPRLVLAARPAALPDGMEILROYAMOLISRASEIRGLA
AVPAGQAGVTLDRDLTAPLSARPAALROIRPILMLCAETLIPDPSIDHLDIDG
QLEPAPGRTEFDWADVTPFSSLAVERAARBYGQVLAIRDAVLEFWTEG
RTAOLRAVSLMRLSALNRAAYGRLEPFGASGASGAATMAAGARCLLELDEPL"
gene
complement(6778..9000)
/ gene="cpeb"
Query Match 3.8%; Score 52.4; DB 2; Length 14922;
Best Local Similarity 46.9%; Pred. No. 0.024;
Matches 275; Conservative 0; Mismatches 296; Indels 15; Gaps 3;
QY 11 ccgaattatgtaagcggtacacactatagtcgaatggaacattgttaata 70
DB 814 CCGGCTGACATCGGTGACGGCGCTGCGAGGCAAGCACTTCGTCTATGC 755
QY 71 gaaggattaaacgacgacgacgacgacgacgacgacgacgacgacgacg 130
DB 754 GCGGGGTCAACACGACGACGACGACGACGACGACGACGACGACGACGAC 695
QY 131 ttgcaataacggtgacacggtccggtatggtatcgtatggtggtgaaatg 190
DB 694 AGGC--GCTGGGCGCAAGCGTCCGGGCTCTCTCGACGGGCGACGCTGA 638
QY 191 aagatgacatcacaagtaagaacattatcttttagcggaagataatttgg 250
DB 637 AGAAGCGCGCGCGGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 578
QY 251 ctgttcttgaagttatgacgacggttatttccattgc-----ttcgctca 301
DB 577 GCGTACTGAGAGTGCACACACGCGGCTACGCGGAGAGCGCGCGCGCGCT 518
QY 302 atctgtctgttatttattgattgaaatgaaatgctttatttgaagaagata 361
DB 517 ACCACGCGCGCGCTACTAGTCGCGCTCAAGAGCTCTCGCGCGCGCGAGAG 458
QY 362 tcattattaatgacgaatgattgttggttggtggaagggatgcttggctgag 421
DB 457 TCATCGTAAATGCGGCAACGACCGCTGGGCAACCGACCGCGCTGGAGCAG 398
QY 422 ggtataaacaagcaatcccgatgctgaaacggtgcttaaacatacttgatg 481

```

```

DB 397 CGACGCTCGGCGGCGTCAAGAACTGCGCGCGCGGCTCCAGCACATCATGTG 338
QY 482 atctcgggggttggaacattcccaatcg---attatgatttgaaagaagtt 538
DB 337 ACCCGCCCACTGGGCGGAGACTGCGAGGCGCTCATGCTGCGCAACGACGCT 278
QY 539 ttatctgacccctcaacgaatacaatttcttgattcataatgta 584
DB 277 ACAGCGCGACCCCGACCGCACTCATCTCTCATCCACATGTA 232
RESULT 13
166494/c 166494 7218 bp DNA PAT 28-DEC-1997
LOCUS Sequence 14 from patent US 5670367.
DEFINITION
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source 1..7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 3.5%; Score 48.6; DB 5; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.19;
Matches 15; Conservative 216; Mismatches 160; Indels 0; Gaps
QY 657 agcattggtgaatttggaacccgtacatacaatgtgtgacgtcatgaagcaag 716
DB 1455 AGAGATGAGAGATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395
QY 717 gacattctgacaagaagaggttggttggtggtggtggtggtggtggtggtg 776
DB 1395 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1325
QY 777 agaattggaatttgacacttcgaattggttggtggtggtggtggtggtggtg 835
DB 1335 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1276
QY 837 aatatacatgtaattgcatatggtttaagaagaacttcgagattgaacacg 896
DB 1275 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1216
QY 897 tacagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 956
DB 1215 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1156
QY 957 tatgcaaggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1016
DB 1155 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1096
QY 1017 ttctaagaagatcatctttaaagcgat 1047
DB 1095 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1065
RESULT 14
AC024591
LOCUS AC024591 111071 bp DNA HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-511G21, WORKING DRAFT
SEQUENCE 13 ordered pieces.
ACCESSION AC024591
VERSION AC024591.3 GI:9256450
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

```


TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 28, 1998 this sequence version replaced g1:1280123.
Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: r.wenham@code.wustl.edu and j.sanger@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is C04D1, 2700 bp overlap; 3' clone is F49E7, 200 bp overlap. Actual start of this clone is at base position 701 of CELK03E6; actual end is at 35049 of CELK03E6

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

1. 35049

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="K03E6"

/chromosome="X"

/complement(3691..7748)

/gene="unc-1"

/complement(join(3691..3766,3931..4145,4822..5095,5154..5278,6432..6545,7296..7326,7735..7748))

/gene="unc-1"

/note="K03E6.5"

/codon_start=1

/product="Erythrocyte band 7 integral membrane protein"

/protein_id="AAC69044.1"

/db_xref="GI:1280125"

/translation="MEYSHQIQPKSCATIDVPDYETIGTIFGVALQSLWIIIVTFPMSVCIVKIVKERYVIFRIGLVFGAGAGGIFIIPIIDYRKIDLRVSYAVP

POELISKDSVYDAVYFRTDIAGVNVDAIYFTKLAQTTLNALGMETIEMTEREATACICETITIDGTEHMGYKVRVVKIRLPOQITRAMAARARAK

VVAEGEKASRALEADAVIQANPVQLQHLQALNSIAEHNSITVFVPEVFGA

FMKRDQ"

/complement(12649..14893)

/gene="K03E6.4"

/complement(join(12649..12750,12801..12928,13397..13527,13570..13666,14217..14391,14837..14893))

/gene="K03E6.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69043.1"

/db_xref="GI:1280126"

/translation="MEKADSVITIRHGNVTRTEPESNVCKLSTVATGDKIINGE

TVSSISFRRANDNIFIDLPFGCLKSSSPFMPRIPLANKPSDKSHKDCVYTGML

PIFTEGKAVCRHCFEPSPSTVOQLIAKIRVNAEIVDPMLPNAGRACRSTIK

GYDLIINVSICSDTAJKFTSTSTIKSSNSHANNSTIADLAGIEAVGNETFN

AOEDNEDESKR"

gene

20796..24499

/gene="11m-6"

/complement(join(20796..20881,21727..21835,22141..22262,23469..23595,23885..24151,24209..24499))

/gene="11m-6"

/note="K03E6.1"

/codon_start=1

/product="similar to other homeobox domains; contains two LIM domains"

/protein_id="AAC69042.1"

/db_xref="GI:1280127"

/translation="MSLLISATSTTEEDKICSGCGLIKRDFEFLIRIKFSKQMA

HILLVDFIITITVLDKSNVMEDESYHESCLRCSCGLISFRCFSGHNGYCEP

DHQIVGRRCRCWTLIPDIYHVRHFMVHQCRCSCGCRPFNGIDYHDEGV

PCRDVDSICNFOETSEPHYITTTPTTISVSNDRPLMEVYRSIHKRTPRRTIN

AQQRQFKIAFERSKSRKRVRLQALRTELISYRVQVWFQNDRAKIKLKNKDSG

DTFRHGGSGRSTEDIRSDDEESYISKLRIGIDIGELMLKIKSVQKRVKFP

"

/complement(25406..26485)

/gene="K03E6.6"

/complement(join(25406..25503,25556..25679,25726..25773,26375..26485))

/gene="K03E6.6"

/note="Similar to Profilin; K03E6.6"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69045.1"

/db_xref="GI:1280128"

/translation="MSMSDIIINNLIGSGNVSKAAILGFDGAWAKSNDNFISVEAV

AGKAFSLDALGTLRLEGKRFVYLNADNDRILKQSGGFIYTIQAVIISIVE

KGLQPEKSTGTGALADYRSIKY"

/complement(28061..33962)

/gene="K03E6.7"

/complement(join(28061..28224,28960..29077,30844..30997,31048..31205,31305..31629,32214..32347,33211..33354,33555..33654,33708..33861,33914..33962))

/gene="K03E6.7"

/note="Similarity to C. elegans protein C01C10.4"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC69046.1"

/db_xref="GI:1280129"

/translation="MDEATSSSEVTEFVSDPTTRQFEEDGHPPLETRHIMIEELE

KINSTDVINKMEVQDLARADRENOVOMSEKLEKQYSSQIAARPEYELKIE

RSLEESQKAAEPERASTISIGIAKOVSITQSTQVSPCLQEVLNHHQIARVE

VEERTRAESLHNSKRAMHLAKIRAMKDNRIATKSRILFEKLETKILEAK

ATILICLAERVAKNDYTSLRNLERISRIHERSTGLESAAVSQEDQKSDFKS

ESLPGNPYPAPAPPEYEDKYIIDKDSDIVIMIKTQDEEERKNSRSLGSGVIL

LAQOLINGNSTEKNHTTPRGSEADISYHTRPVGLSDSDNSEVSLASFNIGDD

TVSKLMSHSELIRKCAAIKRIEIAISVAVAKKRNQRLPRLQVAKRFLRPTPCRKQ

IKNSTLISLDTFLANQSVIFEISAVGERKYPPLSLSECFRRVSGCR"

/complement(5816..6271,611318 t

ORIGIN

BASE COUNT 11644 a 5816 c 6271 g 11318 t

Query Match 3.3%; Score 45.2; DB 33; Length 35049;

Best Local Similarity 51.6%; Pred. No. 1.3; Indels 2; Gaps 1;

Matches 128; Conservative 0; Mismatches 118;

DB 106 caggaactatgcaatgaaaggattgcaaatccggtgtaatacggatcgatg 165

DB 33790 CATGAATATATTCAAAAGACAGAGTTCCAGTGGAGTGCCANCTCTCGAATGTC 33849

DB 166 ttatcgtatggggaacatggaacaaagatgacatcacaagaaagaaacctatcctc 225

DB 33850 TTGTTGTGGGCTCGGAATTCGAATTTTGAAGATTGAAGAAAGATAGT 33909

DB 226 ttaggggaagtaatcattggtgctgcttctgaagttacatgacacggtaaga- 284

DB 33910 TTACCTGAACAATGTTTCGCTGACCTCCAGAGAAGTTGCAATTTCATTTTGT 33965

DB 285 -ttcattgctcgtcctcaatcgtgctgtgattatgattgaatggaagtgccttaa 343

Db 33970 CTTTATAGTTCATCCAAATTTTGTGTGTGTTTGTCTGGAAGTAAACAAATTTTAT 34029

Oy 344 ttggaag 351

|||||

Db 34030 GCGGAATG 34037

Search completed: December 20, 2000, 03:20:08
Job time: 13253 sec

This Page Blank (uspto)

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:58:48 ; Search time 168.09 Seconds

(Without alignments)
3084.148 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470

Perfect score: 1380

Sequence: 1 aatcgaattccgatttta.....ttgataatgattgttagaa 1380

Scoring table: IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq.36.*

| | |
|-----|-------------------------------------|
| 1: | /SID56/gcgdata/geneseq/NA1980.DAT.* |
| 2: | /SID56/gcgdata/geneseq/NA1981.DAT.* |
| 3: | /SID56/gcgdata/geneseq/NA1982.DAT.* |
| 4: | /SID56/gcgdata/geneseq/NA1983.DAT.* |
| 5: | /SID56/gcgdata/geneseq/NA1984.DAT.* |
| 6: | /SID56/gcgdata/geneseq/NA1985.DAT.* |
| 7: | /SID56/gcgdata/geneseq/NA1986.DAT.* |
| 8: | /SID56/gcgdata/geneseq/NA1987.DAT.* |
| 9: | /SID56/gcgdata/geneseq/NA1988.DAT.* |
| 10: | /SID56/gcgdata/geneseq/NA1989.DAT.* |
| 11: | /SID56/gcgdata/geneseq/NA1990.DAT.* |
| 12: | /SID56/gcgdata/geneseq/NA1991.DAT.* |
| 13: | /SID56/gcgdata/geneseq/NA1992.DAT.* |
| 14: | /SID56/gcgdata/geneseq/NA1993.DAT.* |
| 15: | /SID56/gcgdata/geneseq/NA1994.DAT.* |
| 16: | /SID56/gcgdata/geneseq/NA1995.DAT.* |
| 17: | /SID56/gcgdata/geneseq/NA1996.DAT.* |
| 18: | /SID56/gcgdata/geneseq/NA1997.DAT.* |
| 19: | /SID56/gcgdata/geneseq/NA1998.DAT.* |
| 20: | /SID56/gcgdata/geneseq/NA1999.DAT.* |
| 21: | /SID56/gcgdata/geneseq/NA2000.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description | |
|------------|-------|-------------|--------|-------|-------------|--------------------|
| 1 | 1380 | 100.0 | 1470 | 21 | 245335 | DNA encoding a Bac |
| 2 | 897.2 | 65.0 | 1438 | 21 | 245336 | DNA encoding a man |
| 3 | 639.4 | 46.3 | 1482 | 21 | 229845 | Bacillus agaradher |
| 4 | 639.4 | 46.3 | 1482 | 21 | 245337 | DNA encoding a Bac |
| 5 | 601 | 43.6 | 1407 | 21 | 229846 | Bacillus agaradher |
| 6 | 601 | 43.6 | 1407 | 21 | 245338 | DNA encoding a Bac |
| 7 | 408.4 | 29.6 | 1107 | 21 | 245342 | DNA encoding a Bac |
| 8 | 405 | 29.3 | 995 | 21 | 245340 | DNA encoding a Bac |
| 9 | 304.4 | 22.1 | 960 | 21 | 245345 | DNA encoding a Bac |
| 10 | 167.8 | 12.2 | 564 | 21 | 245346 | DNA encoding a par |
| 11 | 126 | 9.1 | 915 | 21 | 245343 | DNA encoding a par |
| 12 | 109.4 | 7.9 | 1461 | 10 | N91621 | Beta-mannase gene |

| | | | | | | |
|----|-------|-----|--------|----|--------|--------------------|
| 13 | 104.4 | 7.6 | 397 | 21 | 245344 | DNA encoding a par |
| 14 | 41.8 | 3.0 | 2635 | 18 | T83966 | DNA encoding a par |
| 15 | 41.8 | 3.0 | 4826 | 18 | V74586 | Staphylococcus aur |
| 16 | 40.4 | 2.9 | 13321 | 18 | V74315 | Staphylococcus aur |
| 17 | 35.4 | 2.6 | 6741 | 21 | A10595 | Gene encoding a su |
| 18 | 35 | 2.5 | 2040 | 17 | T29774 | Bacillus thuringie |
| 19 | 35 | 2.5 | 2100 | 10 | N90712 | cyd gene, Bacill |
| 20 | 35 | 2.5 | 2901 | 12 | O14810 | Btm pg14 72kDa C |
| 21 | 35 | 2.5 | 3558 | 13 | Q22596 | Coding region of |
| 22 | 35 | 2.5 | 9542 | 20 | T50260 | Borrelia burgdorfe |
| 23 | 34.6 | 2.5 | 580073 | 18 | T58840 | Mycoplasma genital |
| 24 | 33.6 | 2.4 | 627 | 20 | Z08233 | Lactobacillus brev |
| 25 | 33.6 | 2.4 | 1395 | 20 | Z08232 | Coding sequence of |
| 26 | 33.6 | 2.4 | 1680 | 20 | Z08234 | Lactobacillus brev |
| 27 | 33.6 | 2.4 | 1763 | 15 | O55316 | Bacillus sp. V230 |
| 28 | 33.4 | 2.4 | 2408 | 19 | V17621 | Btm pg14 72kDa C |
| 29 | 33.4 | 2.4 | 2802 | 12 | O14809 | Staphylococcus aur |
| 30 | 33.2 | 2.4 | 633 | 18 | V75647 | Borrelia burgdorfe |
| 31 | 33 | 2.4 | 782 | 17 | T10931 | orf2 gene sequence |
| 32 | 33 | 2.4 | 1289 | 20 | X26810 | Lactobacillus reut |
| 33 | 33 | 2.4 | 1800 | 20 | Z25012 | Human Th2-specific |
| 34 | 33 | 2.4 | 2703 | 21 | A09056 | Human adenosine re |
| 35 | 32.8 | 2.4 | 6247 | 20 | X86997 | DNA encoding a par |
| 36 | 32.6 | 2.4 | 141589 | 21 | A35005 | Human adenosine re |
| 37 | 32.6 | 2.4 | 141589 | 21 | A35030 | Human adenosine re |
| 38 | 32.6 | 2.4 | 141589 | 21 | A35030 | DNA encoding a par |
| 39 | 32.4 | 2.3 | 1300 | 20 | V82078 | DNA encoding a par |
| 40 | 32.4 | 2.3 | 2132 | 20 | V82077 | Staphylococcus aur |
| 41 | 32.4 | 2.3 | 2841 | 18 | V74488 | DNA encoding human |
| 42 | 32.4 | 2.3 | 3006 | 21 | Z29484 | Salmonella enteric |
| 43 | 32.4 | 2.3 | 22080 | 20 | X06751 | 5' flanking region |
| 44 | 32.2 | 2.3 | 1021 | 16 | T04613 | Listeria phage lys |
| 45 | 32.2 | 2.3 | 1026 | 17 | T13340 | |

ALIGNMENTS

RESULT 1

| ID | Score | Query Match | Length | DB ID | Description | |
|--------|-------------|-------------|--------|-------|-------------|--------------------------------------------------------------------------|
| 245335 | 1380 | 100.0 | 1470 | 21 | 245335 | DNA encoding a Bac |
| XX | | | | | | |
| AC | 245335 | | | | | |
| XX | | | | | | |
| DT | 27-MAR-2000 | | | | | (first entry) |
| XX | | | | | | |
| XX | | | | | | DNA encoding a Bacillus mannanase enzyme. |
| KW | | | | | | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; |
| KW | | | | | | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; |
| KW | | | | | | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; |
| KW | | | | | | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; |
| KW | | | | | | printing paste; plant material degradation; recycled waste paper; |
| KW | | | | | | paper making pulp; guar; locust bean gum; thickener; viscosity; |
| KW | | | | | | mannan-containing food; coffee extract; cleaning composition; |
| KW | | | | | | machine washing; hard-surface cleaner; dishwashing; oral; dental; |
| KW | | | | | | contact lens; body-care composition; fabric softener; oil well drilling; |
| KW | | | | | | subterranean formation fracture; ss. |
| XX | | | | | | |
| OS | | | | | | Bacillus sp. |
| XX | | | | | | |
| FH | | | | | | Key |
| FT | | | | | | CDS |
| FT | | | | | | Location/Qualifiers |
| FT | | | | | | 1..1470 |
| FT | | | | | | /tag- a |
| FT | | | | | | /product- "mannanase" |
| FT | | | | | | /trans_except- (pos: 346..348, aa: Pro) |
| FT | | | | | | /note- "no termination codon given" |
| FT | | | | | | 1..90 |
| FT | | | | | | /tag- b |
| FT | | | | | | 91..1470 |
| FT | | | | | | /tag- c |
| FT | | | | | | /note- "specifically claimed in claim 4; nucleotides |
| FT | | | | | | 91-990 encode the catalytically active core" |

| Db | 1351 | tctctagattatacaaatgttccaacattcttctctaagttaagggaattgtagttcattc | 1410 |
|--------|-------------------------------------------------------------------------|---------------------------------------------------------------|------|
| Qy | 1321 | caatccgcgagtgatgtagtgcgcaaacatcgaattatattgataatgtagtagaa | 1380 |
| Db | 1411 | caatcagcgagtgtagtagtgcgcaaacatcgaattatattgataatgtagtagaa | 1470 |
| RESULT | 2 | | |
| ID | 245336 | | |
| XX | 245336 | standard; DNA; 1438 BP. | |
| AC | 245336; | | |
| XX | 27-MAR-2000 | (first entry) | |
| DE | DNA encoding a mannanase-linker-cellulose binding domain fusion. | | |
| XX | | | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | | |
| KM | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; | | |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; | | |
| KW | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; | | |
| KW | printing paste; plant material degradation; recycled waste paper; | | |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; | | |
| KW | mannan-containing food; coffee extract; cleaning composition; | | |
| KW | machine washing; hard surface cleaner; dishwashing; oral; dental; | | |
| KW | contact lens; body-care composition; fabric softener; oil well drilling | | |
| XX | subterranean formation fracture; cellulose binding domain; ss. | | |
| OS | Synthetic. | | |
| OS | Bacillus sp. | | |
| OS | Clotridium thermoceilum. | | |
| XX | | | |
| XX | Key | Location/Qualifiers | |
| FT | CDS | 1..1431 | |
| XX | | /*tag- a | |
| XX | W09964619-A2. | | |
| XX | 16-DEC-1999. | | |
| XX | | | |
| XX | 10-JUN-1999; | 99WO-DK00314. | |
| XX | | | |
| PR | 10-JUN-1998; | 98US-0111256. | |
| PR | 20-OCT-1998; | 98DK-0001340. | |
| PR | 20-OCT-1998; | 98DK-0001341. | |
| PR | 28-OCT-1998; | 98US-0105970. | |
| PR | 28-OCT-1998; | 98US-0106054. | |
| PR | 23-DEC-1998; | 98DK-0001725. | |
| PR | 05-MAR-1999; | 99DK-0000306. | |
| PR | 05-MAR-1999; | 99DK-0000307. | |
| PR | 05-MAR-1999; | 99DK-0000308. | |
| PR | 09-MAR-1999; | 99DK-0000309. | |
| PR | 10-MAR-1999; | 99US-0123543. | |
| PR | 10-MAR-1999; | 99US-0123623. | |
| PR | 11-MAR-1999; | 99US-0123641. | |
| XX | | 99US-0123642. | |
| PA | (NOVO) | NOVO-NORDISK AS. | |
| XX | | | |
| PI | Kauppinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME; | | |
| XX | WPI; 2000-105891/09. | | |
| DR | P-PSDB; Y54123. | | |
| XX | | | |
| PT | New mannanases for treatment of textiles, plant material and coffee | | |
| XX | extract, and in cleaning compositions | | |
| PS | Example 4; Page 210; 242p; English. | | |
| CC | The present sequence encodes a mannanase-linker-cellulose binding | | |
| CC | domain fusion protein. Mannanase (also known as mannan | | |
| CC | endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase) | | |

| Query Match | 65.0% | Score 897.2 | DB 21 | Length 1438 |
|-----------------------|------------------------------------------------------------------|--------------------|----------|-------------|
| Best Local Similarity | 99.7% | Pred. No. 6.2e-248 | | |
| Matches 899 | Conservative 0 | Mismatches | Indels 0 | Gaps 0 |
| QY 4 | gcaatttcggatatttaagtaagcgtacccctacatcagatgccaatggaaacccattt | 63 | | |
| Db 1 | gcaatttcggatatttaagtaagcgtacccctacatcagatgccaatggaaacccattt | 60 | | |
| QY 64 | gtaatgagaggagatttaacatcgtggcagcagcatggtlaaagccaaagcaactactgcaatt | 123 | | |
| Db 61 | gtaatgagaggagatttaacatcgtggcagcagcatggtlaaagccaaagcaactactgcaatt | 120 | | |
| QY 124 | gaaggagattgcaaataccggtgtctaatacgttcggattgttactcgtatgtyggggacaa | 183 | | |
| Db 121 | gaaggagattgcaaataccggtgtctaatacgttcggattgttactcgtatgtyggggacaa | 180 | | |
| QY 184 | tggaaaaaagttgacatccatacagtaagaacacttactcttttagcggagaatatcat | 243 | | |
| Db 181 | tggaaaaaagttgacatccatacagtaagaacacttactcttttagcggagaatatcat | 240 | | |
| QY 244 | ttagtctgtctctctgtaagttcatgatcctacccggtttatgattccatgcttcgctcat | 305 | | |
| Db 241 | ttagtctgtctctctgtaagttcatgatcctacccggtttatgattccatgcttcgctcat | 300 | | |
| QY 304 | cgtgctgttgattatctgtaattgaaatggaagtgtcttaattggaagaagatacgtc | 363 | | |
| Db 301 | cgtgctgttgattatctgtaattgaaatggaagtgtcttaattggaagaagatacgtc | 360 | | |
| QY 364 | attattaatatttcggaatggaatggtttgtcttcgttggaaaggagatgcttggctgtacggg | 423 | | |
| Db 361 | attattaatatttcggaatggaatggtttgtcttcgttggaaaggagatgcttggctgtacggg | 420 | | |
| QY 424 | tataaacaagcaatcccgcatctggtgaacccgggtctaaacaacataccttgaatgtat | 483 | | |
| Db 421 | tataaacaagcaatcccgcatctggtgaacccgggtctaaacaacataccttgaatgtat | 480 | | |
| QY 484 | gctggtgggttgggacaattccacaacatcgattctgaattatggaagaagaagttttaat | 543 | | |
| Db 481 | gctggtgggttgggacaattccacaacatcgattctgaattatggaagaagaagttttaat | 540 | | |
| QY 544 | gctgaacctcaagaatacaaatgttttcgatctcatatgatatgtaatgtaagtggtat | 603 | | |
| Db 541 | gctgaacctcaagaatacaaatgttttcgatctcatatgatatgtaatgtaagtggtat | 600 | | |
| QY 604 | gcatcgcaagttcgtacataatgtgacccgagttcttaataagaccttcgcatlaagtcatt | 663 | | |
| Db 601 | gcatcgcaagttcgtacataatgtgacccgagttcttaataagaccttcgcatlaagtcatt | 660 | | |
| QY 664 | ggtgaatttggacacgcgtacatacaaatgtgtacgtcgtatggaagcaagattatggtcat | 723 | | |
| Db 661 | ggtgaatttggacacgcgtacatacaaatgtgtacgtcgtatggaagcaagattatggtcat | 720 | | |
| QY 724 | tctgacaacaagaagtggttgggtgttggcgtgtgacatacgtgaagaaggaagcccaaatg | 783 | | |

| Seq | Accession | Gene | Protein | Function | Notes |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|---------|----------|-------|
| Db | 721 | tcgacaaagagagagctgggtggttggtgctgctgcatgtaagggaagcccgcaatg | 780 | | |
| Qy | 784 | gaattattgaacctttcgaatgattggctgcggaataacttaacgttggggaataca | 843 | | |
| Db | 781 | gaattattgaacctttcgaatgattggctgcggaataacttaacgttggggaataca | 840 | | |
| Qy | 844 | atagtcgaatgctcatatgctttaaaggaaccttcgagatgaacacacgtttttacagt | 903 | | |
| Db | 841 | atagtcgaatgctcatatgctttaaaggaaccttcgagatgaacacacgtttttacagct | 900 | | |
| Qy | 904 | gg 905 | | | |
| Db | 901 | ag 902 | | | |
| Result 3 | | | | | |
| ID | 229845 | standard; DNA; 1482 BP. | | | |
| AC | 229845; | | | | |
| DT | 27-MAR-2000 | (first entry) | | | |
| DE | Bacillus agaradherens NCIMB 40482, Mannanase encoding DNA. | | | | |
| KW | Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning; | | | | |
| KW | detergent composition; mid-branched anionic surfactant; washing; | | | | |
| KW | cosmetic stain; food stain; ss. | | | | |
| OS | Bacillus agaradherens. | | | | |
| FT | Key | Location/Qualifiers | | | |
| FT | CDS | 1..1482 | | | |
| FT | | /tag- a | | | |
| FT | | /product- "Mannanase" | | | |
| FT | | /EC number- "3.2.1.78" | | | |
| FT | | /function- "Hydrolysis of 1,4-beta-D-mannosidic linkages" | | | |
| FT | sig_peptide | 1..96 | | | |
| FT | | /tag- b | | | |
| FT | mat_peptide | 97..1029 | | | |
| FT | | /tag- c | | | |
| FT | | /label- Mature_Mannanase | | | |
| PN | W09964552-A1. | | | | |
| PD | 16-DEC-1999. | | | | |
| PF | 10-JUN-1998; | 98WO-US12026. | | | |
| PR | 10-JUN-1998; | 98WO-US12026. | | | |
| PA | (PROC) PROCTER & GAMBLE CO. | | | | |
| PI | Bettol JP, Thoen CAJK; | | | | |
| DR | WPI: 2000-116536/10. | | | | |
| DR | P-PSDB; Y44495. | | | | |
| PT | Detergent composition for removing greasy stains such as cosmetics, | | | | |
| PT | Food stains and body soils | | | | |
| PS | Disclosure; Page 101; 113pp; English. | | | | |
| CC | The present sequence is a DNA encoding B. agaradherens NCIMB 40482, alkaline mannanase enzyme. Mannanase shows maximum activity at pH ranging from 7.5-10.5. It can be used in a detergent composition along with a mid-branched anionic surfactant. The detergent composition may be used for washing purposes, soaking/pre-treatment of stained fabric, hard CC surface cleaning and for removal of cosmetic and/or food stains. This composition provides excellent cleaning effect at low temperature. | | | | |
| SO | Sequence 1482 BP; 446 A; 285 C; 352 G; 399 T; 0 other; | | | | |

| Query Match | Similarity | 46.3% | Score 639.46 | DB 21 | Length 1482 |
|-------------|------------|--------------------------------------------------------------------|---------------------|------------|-------------|
| Best Local | Similarity | 67.5% | Pred. No. 6,76-1174 | | |
| Matches | 933 | Conservative | 0 | Mismatches | 441 |
| | | | | Indels | 9 |
| | | | | Gaps | 2 |
| QY | 4 | gcaaatccggaatttcatgtaagcggtaaccatctatcatgatacgaatgccaatggaaccatctt | 63 | | |
| DB | 94 | gcaagtacagagcttttatgttcatgtaacatcgttatatgatgacgaatgagcgagccattt | 153 | | |
| QY | 64 | gtaatgaaagagatttaaccatcgggcagcagatggtataagaaccaggaactctctgcat | 123 | | |
| DB | 154 | gtcatgaaagagatttaaccatcgagacatctctggtataagaacacggtcttcaacgcat | 213 | | |
| QY | 124 | gaaggagttcgcaaataccggtgctatagctcggtcaggtattgtgttaccgatggggagaca | 183 | | |
| DB | 214 | cctgcacatgccaagcagaagcgccacacacgattctgattgttttccgatgtagcggtcaa | 273 | | |
| QY | 184 | tggacacaaagatgacatccatcaatacagtaagaacacttatctctttagcggaaagatacat | 243 | | |
| DB | 274 | tgggaaaaaagccagacatggaacacacttctgtgaagtcatactgagcttgcggagacaataaa | 333 | | |
| QY | 244 | ttggtgtgtcttcttgaaagttcatatgactcaccgtttagttatgcatctgcttcgctcat | 303 | | |
| DB | 334 | atgggtgctctcgattgaaagttcatatgactcaccgtttagttatgcatctgcttcgctcat | 393 | | |
| QY | 304 | cgtgtctgttgaattatggaattgaaatgaaagtggtgtttaatgtgaaagaaagatacgtc | 363 | | |
| DB | 394 | cgaagcgttgaattatggaatgaaatgaaagtggtgtttaatgtgaaagaaagatacgtc | 453 | | |
| QY | 364 | attataatattgccaagaaatggtgttctgcgtggaagggagatggttggttcgaagg | 423 | | |
| DB | 454 | attataatattgccaagaaatggtgttctgcgtggaagggagatggttggttcgaagg | 513 | | |
| QY | 424 | tataaacaagaacatcccgacttggtgtaacgcgcgtctcaaaacacatccttgatgtgaat | 483 | | |
| DB | 514 | tatatgtgtatcattctccgaagcttcgcgaatgcgcgtctcaaaacacatccttgatgtgaat | 573 | | |
| QY | 484 | gctgcgggtgvgggacaaattccacaatcgaatcgaattatgaaagagaagttttat | 543 | | |
| DB | 574 | gcaacagagatgvgggacaaattccacaatcgaatcgaattatgaaagagaagttttat | 633 | | |
| QY | 544 | gctaaccttcacgaatacaataatgttttcgatttcaatgataatgataagacggtgtaat | 603 | | |
| DB | 634 | gcaagatccggttaaaaaataacagatgttctccatccataatgataatgacggtgtaat | 693 | | |
| QY | 604 | gcatcgcgaagttcgtactaataatgacccgagttcttcaataagaacctcgcattgaatc | 663 | | |
| DB | 694 | gctaacacatggttgaagtcacaaatattgataagattcatagaacacctgtctctgtaata | 753 | | |
| QY | 664 | ggtgaatttggaacacgcgtcatatacaaatgtgtaacgttcggtgaaagcaacgatatgagcat | 723 | | |
| DB | 754 | ggtgaatttcggtcatagacatacgtatggtgtatgcttgaagatacaacatcccttgaat | 813 | | |
| QY | 724 | tctgtaacaagaagagatgtggtgttggttcggtcgtgtcatatgaaagggaaagcccaagatg | 783 | | |
| DB | 814 | tctgtaagaagaactcgcacacgggtgctcgtctgtgttgtaagaaggaatacgaatcgaatg | 873 | | |
| QY | 784 | gagtaatttaagccttctggaatgataatggtgcgtgaaataaacttcaacgcttggggaaatca | 843 | | |
| DB | 874 | gactatttaagccttctcgaagaagactggtcgtgtcaacactttaaactgatttggggaatga | 933 | | |
| QY | 844 | ataaggaatgtgtccataatggtgtttaaagaaacatcagagatttaagacggttttcaagat | 903 | | |
| DB | 934 | attgtccacggggccggaatggtcttaagaagaacatcccaacatccacggtattttcagat | 993 | | |
| QY | 904 | -----ggaggatcgtatgtaagaaactctctccgaacaaactttaaagatttgaagtggt | 957 | | |
| DB | 994 | gataacggtgtgtcaaccctgaaacgcgcaacactgtactactgttgaactttgaagaagac | 1053 | | |
| QY | 958 | atgcaaggaatgactgtgaaagttagcttgagcggagagttccttggtgctgtgacaagatgtgct | 1017 | | |
| DB | 1054 | acacaaaggtgtgcatgtgaaagaaacgttgacgcggttggtcccttgctccgttaacaaagaa | 1113 | | |

OY 1018 tctaaaggaagtcattctttaaagcgatattcaatgtgtcgaattcacacattac 1077
DB 1114 gcttaaggaagtcattctttaaagcgatattcaatgtgtcgaattcacacattac 1173
OY 1078 ttacatgtattcaaatcgtctttacagcgaaatagatagatacactactgttaa 1137
DB 1174 ctgtatagtgaaacaaagtcgtatcttaccgatactctcaactcaacgaacggttcgc 1233
OY 1138 catgcaaatgtgggaagtgtgttaattggaatgcgcgtcttctatgtgaacagga 1197
DB 1234 catgcaaatgtgggaagtgtgttaattggaatgcgcgtcttctatgtgaacagga 1293
OY 1198 catgtttactgttactgtcgtgaagcttgcgcgtatgaacggttc---actggaac 1254
DB 1294 tctgtattacagcagatagcgtcttcttaccgatactcaactcaacgaacgaac 1353
OY 1255 acgcatactctgaattatcaaatgtccaaatcttctcaagtaagggaattggaat 1314
DB 1354 acgcatactcttgaattatcaaatgtccaaatcttctcaagtaagggaattggaat 1413
OY 1315 cagttccaatcagcgagtgatagtgtagacaaacatcgattatattgataatgtatt 1374
DB 1414 caatttcagcgagataatagcagtggttcaactgcctcactaactgtgataactgttact 1473
OY 1375 gta 1377
DB 1474 tta 1476
RESULT 4
ID 245337 standard; DNA; 1482 BP.
AC 245337:
XX 27-MAR-2000 (first entry)
DT
XX
DE DNA encoding a Bacillus mannanase enzyme.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture; ss.
XX
OS Bacillus agaradhaerens.
XX
FH Key Location/Qualifiers
FT CDS 1..1482
FT FT /tag= a
FT FT /product= "mannanase"
FT FT 1..93
FT FT /tag= b
XX
PN MO9964619-A2.
XX
PD 16-DEC-1999.
XX
PE 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1999; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.

PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 98US-0123543.
PR 10-MAR-1999; 98US-0123623.
PR 10-MAR-1999; 98US-0123641.
PR 11-MAR-1999; 98US-0123642.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
XX WPI: 2000-105891/09.
XX P-PSDB: Y54124.
XX
XX New mannanases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions -
XX
XX Example 5; Page 212-213; 242pp; English.
XX
XX The present sequence encodes a Bacillus mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
XX mannanase hydrolyses galactomannans. Specifically, mannanases hydrolyse
XX 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
XX and galactoglucomannans. The mannanase protein, or preparations
XX containing it, are used to improve properties of cellulosic or
XX synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannanases
XX are also used to process coffee extracts (to inhibit gel formation);
XX in cleaning compositions (for machine washing of fabrics, as
XX hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX dental, contact lens or body-care compositions) where they remove
XX mannan-containing soils and prevent binding of some soils to
XX cellulosics; and in fabric softeners. They can also be used in oil
XX well drilling to fracture subterranean formations.
XX
XX Sequence 1482 BP; 446 A; 285 C; 352 G; 399 T; 0 other;
XX
Query Match 46.3%; Score 639.4; DB 21; Length 1482;
Best Local Similarity 67.5%; Pred. No. 6.7e-174;
Matches 933; Conservative 0; Mismatches 441; Indels 9; Gaps 2;
OY 4 gaaatcccgagttatgtatgaagcggtacacatctacatgcagatgcgaatgaacccatt 63
DB 94 gaaatgaagcggtttatgtatgtatgtatgcataatgataatgcgaatgaagcgacatt 153
OY 64 gtaatgaaggaattaacatgagcgacgacatggtataaagaccaggaactactgcaatt 123
DB 154 gtcataagaggtattataacatgacatgcatggttataaagaccatgcacactat 213
OY 124 gaagagattgcgaataacacggtgctatagcgttcggttgcgttcgtatgcgaggaaca 183
DB 214 cctgcacatgcagagcaagcgcccaacgacatgcgtatggtttatgcagatgvgcgcca 273
OY 184 tggacaaaagatgacatccatcagtaagaaacctatctctttagcgaataatcat 243
DB 274 tgggaaaaagacatgacatgcacccatctgtagagcatcttgagcttcgagagaaataa 333
OY 244 ttggttcgtctcttgaagttcaatgacgtctacacggttatgattccattgctgcacat 303
DB 334 atggtgctgttcgttgaagttcaatgacgtctacacggttcggttcgagatattaaat 393
OY 304 cgtgctgttatgtatggaatggaatggaatggaatggaatggaatggaatggaatggaat 363
DB 394 cgaagcgttgatattggaatggaatggaatggaatggaatggaatggaatggaatggaat 453
OY 364 attataatattgcgaatgaatggaatggaatggaatggaatggaatggaatggaatggaat 423
DB 454 attataatattgcgaatgaatggaatggaatggaatggaatggaatggaatggaatggaat 513

```

QY 424 tataacaagaatccgcgattggtacgcgcgtctaaacacattatgtagat 483
DB 514 tatattgtagatccgaaagcttcgcggtctaaacacacattatgtagat 573
QY 484 gctgcggggtgggaacattccacaacatcgatgaattatggagaagatttatt 543
DB 574 gcagagagatggggcaatccacatctatctatgattacgagaagatgtttat 633
QY 544 gctgcacccaacgaataatcttttcgattcatatgttgaattatgcgggtgaat 603
DB 634 gcagatccgttaaaaaatacgaattctccacacatgatatgctggtgtgtat 693
QY 604 gcatcgcaagttcgtaactaatatgacaggttcttaacagaccctcgatgctat 663
DB 694 gctaacactgttagatcaataattgtagagctcagatcagaccctgtctctata 753
QY 664 ggttaatttggacacccgtcacaataatggtgagctcagatgaagacgattatgact 723
DB 754 ggtgaattcggcctacacacacatggtgtgtgtgtgaagatacacatccctatgt 813
QY 724 tctgaacaagaagagattgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 783
DB 814 tctgaagaagaatcgacagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 873
QY 784 gactatttagaccttccgaatgattgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 843
DB 874 gactatttagaccttccgaatgattgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 933
QY 844 atagtgatgtgtcctatgtgttaagaagaacttcgagatgaacccgttttaacagt 903
DB 934 atgttcacagggcgagatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 993
QY 904 -----ggagagatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 957
DB 994 gataacggtgtgtacacccgtgacccgcacatgctacactgtatgacttggagaagc 1053
QY 958 atgcaagagatgagatgagatgactgtgagcgaggtctgtgtgtgtgtgtgtgtgt 1017
DB 1054 acacaaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1113
QY 1018 tctaaaggagatctctttaaagcgagatcattcgtgtgtgtgtgtgtgtgtgtgt 1077
DB 1114 gcttcaggttaacactctttaaagccgagatgtaatttaaccccaattcttcacatgaa 1173
QY 1078 ttcacatgtattcaaaatacgtcttaccagcagaaatagatgaacacgactgtgtaa 1137
DB 1174 ctgtatagatgaacaaagctgttaactacacggtactctcagctcaacgcaacggttcgc 1233
QY 1138 catgcaaatgtgggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1197
DB 1234 catgcaaatgtgggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1293
QY 1198 catggttatacatgtagctgtgaaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1254
DB 1294 tctgattatacatgtagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1353
QY 1255 aagcgtatctctgattatcaaatgtccaacatcttctcaagtaagggaattgagatt 1314
DB 1354 aagcgtatctctgattatcaaatgtccaacatcttctcaagtaagggaattgagatt 1413
QY 1315 aggttcacacacgagatgtagtgaacacacatcgattatgtagatgtatt 1374
DB 1414 caatttcacgagcagatataagcaggtgtcaaacgtctctatagctgtgataacgttact 1473
QY 1375 gta 1377
DB 1474 tta 1476

```

RESULT 5
 229846
 ID Z29846 standard; DNA: 1407 BP.
 XX

```

AC 229846;
XX
XX 27-MAR-2000 (first entry)
DE Bacillus agaradherens Clone MB594, Mannanase encoding DNA.
XX
XX Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing;
XX cosmetic stain; food stain; ss.
OS
XX Bacillus agaradherens.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..1407
XX /tag- a
XX /product- "Mannanase"
XX /EC_number- "3.2.1.78"
XX /function- "Hydrolysis of 1,4-beta-D-mannosidic linkages"
XX sig_peptide 1..93
XX /tag- b
XX mat_peptide 94..1404
XX /tag- c
XX /label- Mature_Mannanase
XX
XX MO9964552-A1.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Bettiol JP, Thoen CAJK.
XX
XX WPI: 2000-116536/10.
XX
XX DR P-PSDB; Y44496.
XX
XX Detergent composition for removing greasy stains such as cosmetics,
XX food stains and body soils -
XX
XX
XX Disclosure; Page 102-103; 113pp; English.
XX
XX The present sequence is the B. agaradherens Clone MB594 DNA encoding
XX alkaline mannanase enzyme. It is derived from B. agaradherens strain
XX NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
XX from 7.5-10.5. It can be used in a detergent composition along with a
XX mid-branched anionic surfactant. The detergent composition may be used
XX for washing purposes, soaking/pre-treatment of stained fabric, hard
XX surface cleaning and for removal of cosmetic and/or food stains. This
XX composition provides excellent cleaning effect at low temperature.
XX
XX Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other;
SQ

```

Query Match 43.6%; Score 601; DB 21; Length 1407;
 Best Local Similarity 67.3%; Pred. No. 6, 9e-163;
 Matches 882; Conservative 0; Mismatches 420; Indels 9; Gaps 2;

```

QY 4 gcaattcggatttatgtgaagcgttacacactctatagatgcaatggaacccatt 63
DB 94 gcaagtacagccttctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 153
QY 64 gtaatgagaggttaacacatggtgcagcagatgtataaagacccagcaactactgaatt 123
DB 154 gtcagagaggtatttaacacatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 213
QY 124 gaaggtattgcaataccggtgtctataacggttcgaggtgtgtgtgtgtgtgtgtgtgtgt 183
DB 214 cctgcattgcagagaagagcgccacacacgattcgtattgtttatcagatggtgcgtcaa 273

```

| | | | |
|----|------|---------------------------------------------------------------------|------|
| Oy | 184 | tgagcaaaagatgacaccccttaagaaagaaaccttatctctttagcggaagataactc | 243 |
| Db | 274 | tggaaaaaagcgcacattgcacacattcgcgaaagctatgagctctgcggagcaaaataaa | 333 |
| Oy | 244 | tgtgtctgtctcttgaaattcatagatgcaccggtttatgattccattcttcgcgtcaat | 303 |
| Db | 334 | atgtgtgcgtctgttgaattcatagatgcacgggtcgcgattccgycggtatattaat | 393 |
| Oy | 304 | cgtgcgtctgattatctgatttgaatgaaatgaaagtgtcttaatttggaaaaggaaacgctc | 363 |
| Db | 394 | cgagcgcgttatctatgataagaatgaaagatgcgttatcgtttaagaagaatacgtct | 453 |
| Oy | 364 | attttaaatattgtaagttaagtgttgcgtctgtgggaagggaatgcttgcgttcagcg | 423 |
| Db | 454 | attatlaacattgtcaaacgcggtgtgtaaaggatttggatgtgctcagcttgggccagatgc | 513 |
| Oy | 424 | tataaacaagaacaccccgcatggttcgtaaacgcgcgtctaaacacatcctctgtgtagat | 483 |
| Db | 514 | tatatgtatgtcatctccgaagcttcgcgatggccggtcttaacacaccccttaattgttat | 573 |
| Oy | 484 | gctgcgggtgtgggacaatttccacaatcgtatcatgatgatgtgaaaggaaagttttaat | 543 |
| Db | 574 | gcagcagagatggggcgaatctccgaatcattcatcatgatattacgcagacagatgtttaat | 633 |
| Oy | 544 | gctgaacctcaacgaataatacagtttcttgatcatatgtatgaaatactcagtggtaat | 603 |
| Db | 634 | gcgaatccggttaaaaaatacagatgttctccatccatataatgatgatattcgttgttat | 693 |
| Oy | 604 | gcacgcgaagctgttaataattgaaacgggtctttaaacaagaacctgcgttatgact | 663 |
| Db | 694 | gctaaacctgttaaaatcaaatatgtaaaggtcatatagatccaagaacctgtctcgttaata | 753 |
| Oy | 664 | ggtaaatcttgacccgcgttcatacaaaatgtgcacgttcgaagcaacgatatatgacat | 723 |
| Db | 754 | ggtaattccggtctcaatagacatactgatgagtgtgatgaaagatacacaactccttatgat | 813 |
| Oy | 724 | tctgtacaagaagaggttgcgtgtgtgtgcgttggtcatctgaaagaagaaacggcccaatg | 783 |
| Db | 814 | tcttaaaagaacctggcacaagaaggtgtcgcgtctgtcttggaaaggcacaagatccagatg | 873 |
| Oy | 784 | gagatattagaaccttcttgaaattgttggtgcctggaaaataaacttaacagcttgggaataca | 843 |
| Db | 874 | gactattatgacctcttcagaagacgtggcctggccaacattcaatcgtatctgggggaatga | 933 |
| Oy | 844 | ataatgaaatgtcccatatgtgtttaaagaagaaccttcgaaattaaacgcggtttttaacgt | 903 |
| Db | 934 | atgttccacggggcgcgttgctttaaacaaggaaacctccaacaacatccacgattattacgat | 993 |
| Oy | 904 | -----ggagatctcgtatgtagagaaacctctccgacaacctcttatgatatttgaagtagt | 957 |
| Db | 994 | gataacggtgtccaccttcgacccgcgacaacgtgcatacctctgatgatctttaagaagagc | 1053 |
| Oy | 958 | atgaaagaagatggactggaanaatagcttgagggggggtcccttgggcgcgttgcaagatgctc | 1017 |
| Db | 1054 | acaacaagggtgcgtcatgtgaagcaaacgcggaacgcgggtgcgttgcgttaacagaatggygt | 1113 |
| Oy | 1018 | tctaaagaagatcattcttttaaaagcgaattacaatttgcgtcaaatccacaatactac | 1077 |
| Db | 1114 | gcttcagtgtaactactctctttaaaagccgaatgataatttaactcaaatcttccacaatga | 1173 |
| Oy | 1078 | ttaacagtattcaaaatacgccttcaacagacgaatagatagatacaaacgatacgtttaa | 1137 |
| Db | 1174 | ctgtatagtgaaacaaagctgtaatctacacggaatacttcacgtccaacgcacgcgttcgc | 1233 |
| Oy | 1138 | catgcgaattggggaagtgtgtgttaatggaatgacgcgcgtcttatatgttgaaacagga | 1197 |
| Db | 1234 | catgcgaattggggaatctccggttaatggaatgaaatgcaagaactttaaactgtgaaaacgggc | 1293 |
| Oy | 1198 | catggttatacatgctactctcggaaagctttgtgcgattaaacggttc---actctgaaca | 1254 |
| Db | 1294 | ctgtattataaattgacctagcgttccctctttaacgcgtatacaatagctcccaactccagaaaca | 1355 |
| Oy | 1255 | acgcatactctagattatcatcaaatgtccaaaactctttctcaagttaaggga | 1305 |

| Seq ID | Seq | Length | GC | GC3 | GC4 | GC5 | GC6 | GC7 | GC8 | GC9 | GC10 | GC11 | GC12 | GC13 | GC14 | GC15 | GC16 | GC17 | GC18 | GC19 | GC20 | GC21 | GC22 | GC23 | GC24 | GC25 | GC26 | GC27 | GC28 | GC29 | GC30 | GC31 | GC32 | GC33 | GC34 | GC35 | GC36 | GC37 | GC38 | GC39 | GC40 | GC41 | GC42 | GC43 | GC44 | GC45 | GC46 | GC47 | GC48 | GC49 | GC50 | GC51 | GC52 | GC53 | GC54 | GC55 | GC56 | GC57 | GC58 | GC59 | GC60 | GC61 | GC62 | GC63 | GC64 | GC65 | GC66 | GC67 | GC68 | GC69 | GC70 | GC71 | GC72 | GC73 | GC74 | GC75 | GC76 | GC77 | GC78 | GC79 | GC80 | GC81 | GC82 | GC83 | GC84 | GC85 | GC86 | GC87 | GC88 | GC89 | GC90 | GC91 | GC92 | GC93 | GC94 | GC95 | GC96 | GC97 | GC98 | GC99 | GC100 | GC101 | GC102 | GC103 | GC104 | GC105 | GC106 | GC107 | GC108 | GC109 | GC110 | GC111 | GC112 | GC113 | GC114 | GC115 | GC116 | GC117 | GC118 | GC119 | GC120 | GC121 | GC122 | GC123 | GC124 | GC125 | GC126 | GC127 | GC128 | GC129 | GC130 | GC131 | GC132 | GC133 | GC134 | GC135 | GC136 | GC137 | GC138 | GC139 | GC140 | GC141 | GC142 | GC143 | GC144 | GC145 | GC146 | GC147 | GC148 | GC149 | GC150 | GC151 | GC152 | GC153 | GC154 | GC155 | GC156 | GC157 | GC158 | GC159 | GC160 | GC161 | GC162 | GC163 | GC164 | GC165 | GC166 | GC167 | GC168 | GC169 | GC170 | GC171 | GC172 | GC173 | GC174 | GC175 | GC176 | GC177 | GC178 | GC179 | GC180 | GC181 | GC182 | GC183 | GC184 | GC185 | GC186 | GC187 | GC188 | GC189 | GC190 | GC191 | GC192 | GC193 | GC194 | GC195 | GC196 | GC197 | GC198 | GC199 | GC200 | GC201 | GC202 | GC203 | GC204 | GC205 | GC206 | GC207 | GC208 | GC209 | GC210 | GC211 | GC212 | GC213 | GC214 | GC215 | GC216 | GC217 | GC218 | GC219 | GC220 | GC221 | GC222 | GC223 | GC224 | GC225 | GC226 | GC227 | GC228 | GC229 | GC230 | GC231 | GC232 | GC233 | GC234 | GC235 | GC236 | GC237 | GC238 | GC239 | GC240 | GC241 | GC242 | GC243 | GC244 | GC245 | GC246 | GC247 | GC248 | GC249 | GC250 | GC251 | GC252 | GC253 | GC254 | GC255 | GC256 | GC257 | GC258 | GC259 | GC260 | GC261 | GC262 | GC263 | GC264 | GC265 | GC266 | GC267 | GC268 | GC269 | GC270 | GC271 | GC272 | GC273 | GC274 | GC275 | GC276 | GC277 | GC278 | GC279 | GC280 | GC281 | GC282 | GC283 | GC284 | GC285 | GC286 | GC287 | GC288 | GC289 | GC290 | GC291 | GC292 | GC293 | GC294 | GC295 | GC296 | GC297 | GC298 | GC299 | GC300 | GC301 | GC302 | GC303 | GC304 | GC305 | GC306 | GC307 | GC308 | GC309 | GC310 | GC311 | GC312 | GC313 | GC314 | GC315 | GC316 | GC317 | GC318 | GC319 | GC320 | GC321 | GC322 | GC323 | GC324 | GC325 | GC326 | GC327 | GC328 | GC329 | GC330 | GC331 | GC332 | GC333 | GC334 | GC335 | GC336 | GC337 | GC338 | GC339 | GC340 | GC341 | GC342 | GC343 | GC344 | GC345 | GC346 | GC347 | GC348 | GC349 | GC350 | GC351 | GC352 | GC353 | GC354 | GC355 | GC356 | GC357 | GC358 | GC359 | GC360 | GC361 | GC362 | GC363 | GC364 | GC365 | GC366 | GC367 | GC368 | GC369 | GC370 | GC371 | GC372 | GC373 | GC374 | GC375 | GC376 | GC377 | GC378 | GC379 | GC380 | GC381 | GC382 | GC383 | GC384 | GC385 | GC386 | GC387 | GC388 | GC389 | GC390 | GC391 | GC392 | GC393 | GC394 | GC395 | GC396 | GC397 | GC398 | GC399 | GC400 | GC401 | GC402 | GC403 | GC404 | GC405 | GC406 | GC407 | GC408 | GC409 | GC410 | GC411 | GC412 | GC413 | GC414 | GC415 | GC416 | GC417 | GC418 | GC419 | GC420 | GC421 | GC422 | GC423 | GC424 | GC425 | GC426 | GC427 | GC428 | GC429 | GC430 | GC431 | GC432 | GC433 | GC434 | GC435 | GC436 | GC437 | GC438 | GC439 | GC440 | GC441 | GC442 | GC443 | GC444 | GC445 | GC446 | GC447 | GC448 | GC449 | GC450 | GC451 | GC452 | GC453 | GC454 | GC455 | GC456 | GC457 | GC458 | GC459 | GC460 | GC461 | GC462 | GC463 | GC464 | GC46 |
|--------|-----|--------|----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
|--------|-----|--------|----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|

preparations containing it, are used to improve properties of cellulosic or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based sites or printing pastes). They are also used to degrade or modify plant materials (particularly recycled waste paper, paper making pulps, or material containing guar or locust bean gums (thickeners), or to reduce viscosity of mannan-containing foods or feeds). The mannanses are also used to process coffee extracts (to inhibit gel formation); in cleaning compositions (for machine washing of fabrics, as hard-surface cleaners, for hand or machine dishwashing, also in oral, dental, contact lens or body-care compositions) where they remove mannan-containing soils and prevent binding of some soils to cellulosics; and in fabric softeners. They can also be used in oil well drilling to fracture subterranean formations.

Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other;

| | | | | |
|---------------------------|-------|---------------------|-----------|--------------|
| Query Match | 43.6% | Score 601; | DB 21; | Length 1407; |
| Best Local Similarity | 67.3% | Pred. No. 6.9e-163; | | |
| Matches 882; Conservative | | 0; Mismatches 420; | Indels 9; | Gaps 2 |

| | | | |
|----|-----|------------------------------------------------------------------------|-----|
| QY | 4 | gcattccggaatttatgtatgaacggttacacatctcatacgaatgcnaatgcnaaccattc | 63 |
| Db | 94 | gcaagtaacagaccttctatgtctgataaggcaataaagatctatctgcacgaacaaatgcgaaccattc | 153 |
| QY | 64 | gtaatgagagaggtatcaaccatgagccacgaatcgtataaagacacgaagcgaacactgcgcattc | 123 |
| Db | 154 | gtcatgagaggtatcaaccatgagacctgtgtgtataaagacacgaaccttcaacagacctattc | 213 |
| QY | 124 | gaagggattgcgaataacacggtgtcctaatacgcgtccgaattggtctatcgtatgvtggtgagca | 183 |
| Db | 214 | ccctgcacattgcagagacagagcgcacaaacgaatctgattgttttatacgaatgvggtgcac | 273 |
| QY | 184 | tggacaagaagatgatacccatcacagtaagaagaacctatctcttcaagcggagaatacatc | 243 |
| Db | 274 | tggagaanaaagacaacatctgcacacatctcgtatgaagatgcatactgcgttcgtgcgacaaataaa | 333 |
| QY | 244 | ttagtgcgtctctcgaagatcatgatacgtgtacacggttatgatacttcacattgcttcgcatac | 303 |
| Db | 334 | atggtgcgtcgtctgtagagctcatagatgacaaaggttcgcgaattccgcgaatgattaaat | 393 |
| QY | 304 | cgtgcctgtgattatctgatactgaaatgagaaatgaccttaattgtaaaagaaagataccgcgc | 363 |
| Db | 394 | cgaagccggtgattatgtagatagaatagaaagatgcgtcttaacgtataaagaagatacagctt | 453 |
| QY | 364 | attattaatatctgcgaatgaatgaggttctgttcgvggagagaggtatgctgtgcgcagcg | 423 |
| Db | 454 | attataaatatctgcgaacagagcgtgcataagagatctgvggattggtctcaagcttgggcgcgaatgc | 513 |
| QY | 424 | tataaacaagacaaaccgcggaattgtagtaacgcgcgtcccaaaacacataccatgcgtatgcagat | 483 |
| Db | 514 | tatatatgctcatctcccggaagcttcgcgaatgcgcgtcccttaaacacacacttaagtttgc | 573 |
| QY | 484 | gctgcgvggttgcgaatctccacaaatcgaattcagatactgatactgaaagaaagattttaat | 543 |
| Db | 574 | gcagcaggtatgvggacaaatctcgaaatctcttcaatgattaaagacaagaatgtgtttaat | 633 |
| QY | 544 | gctgaaccccaacggaatacaaatgttttcgattatatagtatgaataatgaaagtgttaat | 603 |
| Db | 634 | gcgaagatccgtataaaaaataacgatgtcttcctccaccatactgcagatagatgctgtgtgtat | 693 |
| QY | 604 | gcatacgaaagtcgtactaataatctgacacgaattcttaatacagaacccctgcataatgacatc | 663 |
| Db | 694 | gctaacaacgttatgatacaatatattgataagatctgataagacacacttgcctcgttaata | 753 |
| QY | 664 | ggttgaaattgcgcacacggtccatacaaaatgvtgacgtgcgaaggaaacgaatctatgagctat | 723 |
| Db | 754 | ggatgaattcgtgcataagacatactgatactgatactgataagatacaaatcccttagttat | 813 |
| QY | 724 | tctgaaacaaagagvggttgggttgggtgcgtgcataaggaagagaaagcgcgaagatg | 783 |
| Db | 814 | tctgaaagaaactgcgcacaggtgcgtccgcttgggtccttggaagagcaacagctatccgaattg | 873 |

| | | | | | | |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|--|------|---------------------------------------------------------------------|------|
| OY | | | | 784 | gagtatattgaaccttccttgaaatgatgtggcctcggaataacacttaacagccttggggaataca | 843 |
| Dd | | | | 874 | gaactatttagacctttccagaagactgycctgtgcacaattaactatgtatgggggaataga | 933 |
| OY | | | | 844 | aatagttaaaggccaatatagtgtttaagagaaaccttgagatgaatgaaccgttttaacagt | 903 |
| Dd | | | | 934 | attgttcaaaggggccgcatgcttaacagaagaacctccaacatcacaccgatatttacagat | 993 |
| OY | | | | 904 | -----ggaagatcgtatgtaggaagaaaccttcoccgaaaacactttaatttgtgaagtagt | 957 |
| Dd | | | | 994 | gataaaggtgtgcacccttgaaacccgcaactgtactactcttgatgactttgtaggaaggaac | 1053 |
| OY | | | | 958 | atgcgaaggttgagacgtgggaagtagctctgsgcggaagctccttggtctgtgacagaagtgtct | 1017 |
| Dd | | | | 1054 | acacaagggctgcataggaagcaacgtlgaccggttgcccttgttcgttaacagaaatgggt | 1113 |
| OY | | | | 1018 | tctaagaagaaagcatcttctaanaagcggatatctcaattgtcgtccaattcacacattac | 1077 |
| Dd | | | | 1114 | gtcttcaggaactactcttctaanaagccgcatgtgtaatttaactcaaatcttcacagaa | 1173 |
| OY | | | | 1078 | ttaactgtattccaataatcagctccttacagcagaatgtagatacaagctactgtitaa | 1137 |
| Dd | | | | 1174 | cgtatagtgaaacaagaacgtgaattacacagataacttcagctcaacgcaacgcttcgc | 1233 |
| OY | | | | 1138 | catgcaaanittgggggaagtggttgtgtaatggaatgtagctgcgctctttagtgaanaagga | 1197 |
| Dd | | | | 1234 | catgcgaatttgggggaatcccgcgtaatgtcgtatgtaatgcaagacttcacgtgaaaaagcgc | 1293 |
| OY | | | | 1198 | catggtttacaatgtgtactctgtgaagcttltgtccgatttaacygttc--atctgacaa | 1254 |
| Dd | | | | 1294 | tcgtattatcatactgcatagcgcgtccctttacacgtatcalcatagctccaactcagacaa | 1353 |
| OY | | | | 1255 | acgcatctcagatattatcaaatgtyccaanatcttctcaagaagaagaa | 1305 |
| Dd | | | | 1354 | acgtattcttttgaatttaacaacacatcogaataatcatcatgtttaaggaaa | 1404 |
| RESULT | 7 | | | | | |
| ID | Z45342 | | | | | |
| XX | Z45342 standard; DNA; 1107 BP. | | | | | |
| AC | Z45342; | | | | | |
| DT | 27-MAR-2000 (first entry) | | | | | |
| DE | DNA encoding a Bacillus sp. AA349 mannase enzyme. | | | | | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; galactoglucomanman; celulosic fib-re synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; plant material degradation; recycled waste paper; paper making pulp; guar; locust bean gum; mannan-containing food; coffee extract; cleaning composition; machine washing; hard-surface cleaner; dishwashing; oral; dental; contact lens; body-care composition; fabric softener; oil well drilling; subterranean formation fracture; ss. | | | | | |
| XX | Bacillus sp. | | | | | |
| OS | | | | | | |
| FH | Key | Location/Qualifiers | | | | |
| FT | CDS | : 1..1107 | | | | |
| FT | | : /*tag= a | | | | |
| FT | | /product= "mannanase" | | | | |
| FT | sig_peptide | /note= "not termination codon given" | | | | |
| FT | | : 1..75 | | | | |
| FT | | : /*tag= b | | | | |
| XX | | | | | | |
| PN | MO9664619-A2. | | | | | |
| XX | | | | | | |
| PD | 16-DEC-1999. | | | | | |
| XX | | | | | | |
| PF | 10-JUN-1999: | 99WO-DK00314. | | | | |
| XX | | | | | | |

PD 16-DEC-1999 .
XX
PF 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001275.
PR 03-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
DR WPI: 2000-105891/09.
DR P-PSDB: Y54127.
XX
PT New mannanases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions
XX
PS Example 10; Page 219-220; 242pp; English.
XX
CC The present sequence encodes a *Bacillus* mannanase (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannanases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannan, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (nonwoven) fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.
XX
SQ Sequence 995 BP; 286 A; 162 C; 256 G; 291 T; 0 other;

| | Query Match | Similarity | 29.3% | Score 405 | DB 21 | length 995 |
|----|-------------|--------------------------------------------------------------------|--------------|-------------|------------|------------|
| | Best Local | Similarity | 66.1% | Pred. No. 1 | le-106 | |
| | Matches | 585 | Conservative | 0 | Mismatches | 300 |
| | | | | | Indels | 0 |
| | | | | | Gaps | 0 |
| QY | 13 | ggatttatgtgaagcgtgtccactctctatcgcgtacgtacgtgaacccatttgaatgaga | 72 | | | |
| DB | 106 | gggtttccatgtttatgtgtacacattatgtatgtgcaaatgtgaacccctttgtatgaga | 165 | | | |
| QY | 73 | gggattaacccatgtgggacgcacgtgtataaagccagccacactctgcaattgaaggatt | 132 | | | |
| DB | 166 | gggatttaacatcgtgacatctgttgtttaacaagaacatgaaacacatcattgagaaggatt | 225 | | | |
| QY | 133 | gcaaataccggtggtcraatcgggtccggatctgtgtatcgaatggggagaaatgagaaaa | 192 | | | |
| DB | 226 | agtcacaacggggccaatcgcattctgtgcgttttgcataatgtggcaaaagtgcacaaaa | 285 | | | |
| QY | 193 | gatgacatcccatcacaagtgaagaaaccttactctcttgcggagaagatatcatcttggttgc | 252 | | | |
| DB | 286 | gatgatcgaaaacatgtgtactcttcgtttattctcttgcgcagagacagcatcaaatgtatggcc | 345 | | | |

| QY | 253 | gctcttggaattcatatgctacacggcttcttgatccatctgctgcgaacacgctgcgtt | 312 |
|--------|-------------|----------------------------------------------------------------------------|-----|
| Db | 346 | gcttttagaagttcatatgactactactgtagcacaataattctctcgatctgcgaagcgcgcgtg | 405 |
| QY | 313 | gattatggaattgaaatgsgaagtcgtcttaatttgsgaaggaagataccgctatattaat | 372 |
| Db | 406 | gactatctgattgagatgtagaaggatttcttcgcgggggaaagagaacatagtatcataat | 465 |
| QY | 373 | attgcgagatgaagcttgcttcgtctgsgagaggsgatgctctggtgcgcgaaggtataaacaa | 432 |
| Db | 466 | atcgccaatgaatggtacggtgctcttgsgacggagcgcatgsgcgcagsggtatccagat | 525 |
| QY | 433 | gcacatccggaattgcgtacacgccgggtctaaacacataccctgatgatgtagctgcgggg | 492 |
| Db | 526 | gcgatacgctacgcttcgaaatgcaggtctgttcacatacatattatggttagcgtgcggt | 585 |
| QY | 493 | tggsggaataattccacaatcgtatctatgatttttgsgaaggaagatttttaatgctgcaacct | 552 |
| Db | 586 | tatggccagatccctaatctcggtagttgatttcttgccaagaagatataaagctgcgaac | 645 |
| QY | 553 | caacggaatacaatgcttttcgattcatatgtagtaataatgcaggtgtagatgcacgcga | 612 |
| Db | 646 | cagsgaatacaacaaatgctttctcgttcatatgtagtaataatgcagsgcggagatgcataca | 705 |
| QY | 613 | gttcgtctaatttttaccgagttcttctaatacgaacctgcattagctatggtggaatt | 672 |
| Db | 706 | gttagaagaaacattgactcgaatcttaagccgaacattagctctgtcatgttgatattc | 765 |
| QY | 673 | ggacacgcgtatatacaaatggtgcgcgtgcgaatgaagcaacgattatgactatgaaacaa | 732 |
| Db | 766 | ggcgattggcattatgacagcgtagtctgttgtaggacacacatttaagctatccagaacaa | 825 |
| QY | 733 | agagagagctgggtggtgctggtgcgtgctcagtgaaaggggaacgcccagaatggagattctta | 792 |
| Db | 826 | agaatatgtgagatggtgtgcgtgtagaactgagcatalgtagaaggggtgcgaataattct | 885 |
| QY | 793 | gaaccttcgagatgattgsgctgtagaataacacctacagcgtctgsgaataacataagtagat | 852 |
| Db | 886 | gattatccgaatggaacttgcctgctgtagaactgcacctgacatggtgggtgcgacgaaatgaaac | 945 |
| QY | 853 | ggtccatagttttaagagaagaacttcgagattaaagcaccgctttt | 897 |
| Db | 946 | ggtccgaatggattcgtaacacctcaaaagaagagcagtggttt | 990 |
| RESULT | 9 | | |
| ID | 245345 | standard; DNA; 960 bp. | |
| XX | 245345; | | |
| XX | 27-MAR-2000 | (first entry) | |
| DE | | DNA encoding a <i>Bacillus</i> sp. mannannase enzyme. | |
| XX | | | |
| KW | | Mannannase: mannan endo-1,4-beta-mannosidase; beta-mannanase; | |
| KW | | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | |
| KW | | mannan; galactomannan; glucomannan; galactoglucomannan; celluloseic fibre; | |
| KW | | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | |
| KW | | plant; material degradation; recycled waste paper; paper making pulp; | |
| KW | | guar; locust bean gum; mannan-containing food; coffee extract; | |
| KW | | cleaning composition; machine washing; hard-surface cleaner; | |
| KW | | dishwashing; oral; dental; contact lens; body-care composition; | |
| KW | | fabric softener; oil well drilling; subterranean formation fracture; ss. | |
| OS | | <i>Bacillus</i> sp. | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | CDS | 1..960 | |
| FT | | /tag- a | |
| FT | | /product- "mannanase" | |
| FT | | /note- "not termination codon given" | |
| FT | sig_peptide | 1..84 | |

CC The present sequence encodes a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannanases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.

Oy 3 tgcgaattccggaatttttgcgaacggttaacactctatagaatgccaattgaaccatt 62
 81 tgcgaataagcggtcttcacgtaaaaggtacagattttgcgaacaaatgcgatcctta 140
 Oy 63 tgtaatgtgaaaggatttaaccctatggcagcagcatggtatataaagaccaggaactctgcaat 122
 Db 141 cgtctacgctgcgccaacccctgacatctctctgttttaaaccaagattctagaagagccaat 200
 Oy 123 tgaaggaggttcgcaaataccggtgcataatagcttcggatattggttatctgatggggaca 182
 Db 201 cccctccctatagcagaataacaggggcgaaacacgctgaagatcgtcttatccatgtgaagcc 260

| QY | 103 | atggaacaaagatgacatccataacgaataaagaacactatctcttttgcggaagatatca | 242 |
|-----------|--------------------------------------------------------------------------|---------------------------------------------------------------------------------|-----|
| Db | 261 | atgggaaacaaagatgacatccctctgagctgcccgtgtgtctgtgcacacagaacacatg <th>320</th> | 320 |
| OY | 243 | ttgtgtgctcttcttgaaagtcatatgtctacccgggtatgatcttcattgcttcgtctca <th>302</th> | 302 |
| Db | 321 | gttgcaaacctgtctggaagtccacgagtctacaggaagtgtacaatcccgatattaa <th>380</th> | 380 |
| OY | 303 | tcgtgtgttattatttgattgattgaataagaaagtgtcttaatttggaagaaagataacgt <th>362</th> | 362 |
| Db | 381 | taagcagtcgattacgtgatcgtgaatgtgcgtatgtcttaagggagcagaagaccggt <th>440</th> | 440 |
| OY | 363 | cattataatcttgcgaatgaatggtttgtgttcgttgggaaggagatgcttgggtcag <th>422</th> | 422 |
| Db | 441 | aatccattacacattgcgaatgatgtatgtgggcgttggagagtgtacggttttggcagaag <th>500</th> | 500 |
| OY | 423 | gtataaagaacaaatcccgagattgctgaacgcgggtctaaacataccttgtgttaa <th>482</th> | 482 |
| Db | 501 | atacgacaagaagatcccgctgtgcgagctgtcggccttcgccacatacgttaatagtga <th>560</th> | 560 |
| OY | 483 | tgcttcgggggggggacaaatttccacaacatcgatcatcgaattatggaagaagaattttaa <th>542</th> | 542 |
| Db | 561 | tgcgcaaggtgtggggacacgaacccctgctctatccacagaacgggggacgacatttgc <th>620</th> | 620 |
| OY | 543 | tgtctgacccccaacgaataacataatgttttcgattcatatgtatgaatgtcaggtgtaa <th>602</th> | 602 |
| Db | 621 | ctccgattccattaaaaaacaacaaatgtttccattccataatgtacgaatatgcagaagcga <th>680</th> | 680 |
| OY | 603 | tgcacatcgaaatttcgtactaatatcttgaccgaagtcttaatacgaactcgcattagtc <th>662</th> | 662 |
| Db | 681 | tagggcgacagttctctgaanaacatcgacgggtgtactctgcgaataatcttctgtgtaat <th>740</th> | 740 |
| OY | 663 | tgtgtgaatttggacacccgtcatcaaaatgtgtgaactgcgtgaagcaagattatgaccta <th>722</th> | 722 |
| Db | 741 | cgtgtgaatttggccatagatgacatcatctatgttcgagatgcgtgaagaatgtcgatttgccta <th>800</th> | 800 |
| OY | 723 | tctctgaacaaagaggagttgggtgtgttggcgtgtgtcatgtgaaaggggaacggcccaaatg <th>782</th> | 782 |
| Db | 801 | tacagcagagagggaagttgggtgtgtgtgtgtgtgtatgtatgtacatagtcgggggtgt <th>860</th> | 860 |
| OY | 783 | ggagattatagaaccttcogaatgatgtggctgtgaaataacctaacgcttggggaataac <th>842</th> | 842 |
| Db | 861 | tgaactacttga---tttaactgaagggccatcgaatcattacgaagtgtggggcgaacg <th>917</th> | 917 |
| OY | 843 | aatagtgaatgtccatagtttaagaagaact <th>876</th> | 876 |
| Db | 918 | gattgtctatgtgggaatgggcttaagaagaact <th>951</th> | 951 |
| RESULT 10 | | | |
| 245346 | | | |
| ID | 245346 standard; DNM; 564 BP. | | |
| XX | XX | | |
| XX | XX | | |
| XX | XX | | |
| DT | 27-MAR-2000 (first entry) | | |
| DE | DNA encoding a partial <i>Bacillus</i> sp. mannanase enzyme. | | |
| XX | XX | | |
| XX | XX | | |
| MM | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | | |
| KW | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | | |
| KW | mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre | | |
| KW | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | | |
| KW | plant material degradation; recycled waste paper; paper making pulp; | | |
| KW | guar; locust bean gum; mannan-containing food; coffee extract; | | |
| KW | cleaning composition; machine washing; hard-surface cleaner; | | |
| KW | dishwashing; oral; dental; contact lens; body-care composition; | | |
| KW | fabric softener; oil well drilling; subterranean formation fracture; ss. | | |
| XX | XX | | |
| OS | <i>Bacillus</i> sp. | | |
| XX | XX | | |
| EH | Key | | |
| FT | Location/Qualifiers | | |
| CDS | 1..564 | | |

PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX WPI: 2000-105891/09.
 DR P-PSDB: Y54130.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 XX extract, and in cleaning compositions
 PS Disclosure: Page 225; 242pp; English.
 XX
 CC The present sequence encodes a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to celluloses; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 SQ Sequence 915 BP; 267 A; 203 C; 229 G; 216 T; 0 other;

Query Match 9.1%; Score 126; DB 21; Length 915;
 Best Local Similarity 51.9%; Pred. No. 1.4e-26;

Matches 493; Conservative 0; Mismatches 400; Indels 57; Gaps 7;

OY 436 atccgcgagtcgttaacccgcgttcaaacatcttgatgtgtaacgtcgggggtg 495
 DB 1 atctcaacatctgaagaatcgcgttctgcatacaacatcgttctggtgagtcacgcggggtg 60
 OY 496 ggaacattcccaacatcgttcatgataltatggaagaaagttttatgtaccctca 555
 DB 61 gacacaatctatcgcaatlaagcttatggaagaagtggttaacatgatccgcg 120
 OY 556 cgaatataatgttttgatcatatgtaataatgacagtggttaatgcatcgcaagt 615
 DB 121 cgaatgtatgtcttcatacaatgtagcgttctgttgataatagtcgcgaatc 177
 OY 616 cgtactaatatgacgcgagttcttaagaacctgcgttgatgtgtaattgta 675
 DB 178 ggcagcgaatcgaagccatcaagaacctgtgtcgcgtcaatgcttggtgaatc 237
 OY 676 caccgtcatacaatg-----gtgaagtcgatatgaacacgattag 717
 DB 238 tactaactacaacacgcaataaacaactgtggagtcaggttaacgcccagaatatac 297
 OY 718 agcttcttgacaagaagaggttggtgtgtgcgtgtcatggaaggaagccca 777
 DB 298 aatcagcgcaagcaaaaagaaatcggtcatcgcgtgtgtcgtgactggaatgacgcg 357
 OY 778 gaatggagatattgaccttcgaatgttggtgcgtgaataaacttcagacttgagga 837
 DB 358 gtaactctctgtgtgattgacacaaa---cgattgcaaacacttcacatcattg 414
 OY 838 aatacaatagtgaatggtcatatggtttaagaaacttcagatgaacacggtttt 897
 DB 415 aatctagtgttaatgaaacacagcgcatcgtgactgctgtccacgaactgtattt 474

OY 898 acaagtgagagatctgattgaggaacttctccgacaactcttatgatttga---agt 954
 DB 475 -----aatacaacaacaacttatgatttgaagcggc 510
 OY 955 agtatgaagagtagctggaagtagcttgagcggaggtccttggtgcaagtg 1014
 DB 511 aatgccagaggtgttcaggttcggtttgagcgggggacctgtgttattatg 570
 OY 1015 tctttaaaggaagtcattctttaaagcggaatattcaatgtcgtcaaatcacaat 1074
 DB 571 ggcgcgagcggtatgtattctctcaagcgaatatactcttagcgcca---ctcaaaa 627
 OY 1075 tactataatgttatccaataatcgtcttaacgcggaatagatgataaactactgt 1134
 DB 628 gcttgcacaacacagcgctccataattcagcgccggtctacattccgtaaga 667
 OY 1135 aaacatgcaattggggaagtggttgtaaggaatgacgtcgcgtcttatgtaaa 1194
 DB 688 aagcatgcagcttggtggaatcaacgcgcggtatgcaagccaagtattatgtaaa 747
 OY 1195 ggaatggttatatacatggtactcttgaaagcttgcggaatgaagttcatctg 1254
 DB 748 ggggcggttaacgctggtatgagcgcgactgtaaacatcaacacagctcg---ggcaac 804
 OY 1255 acgtatctctgattatcaaatgtccaatcttctcaagtaaggaaattgagtt 1313
 DB 805 aatgtacgcttaaacctgcgagcatctcctaactgaacgactgaagaatcggaa 864
 OY 1315 cagttccaatcagcgagtagtagtgacaaacatcgattatga 1364
 DB 865 gaattataacacccgcaaatcgggtgttcttcgcgaattatggtga 914

RESULT 12

ID N91621 standard; DNA: 1461 BP.

XX AC N91621;

DT 13-MAR-1990 (first entry)

DE Beta-mannase gene of alkalophilic *Bacillus*.

XX KM *Escherichia coli*.

XX OS *Bacillus* sp.

XX FH Key Location/Qualifiers

FT CDS 1..1461

PN JP01228477-A. /tag= a

XX PD 12-SEP-1989.

XX PF 09-MAR-1988; 88JP-0053774.

XX PR 03-MAR-1988; 98JP-0053774.

XX PA (SHKJ) SHINGIJUTSU KAIHATSU.

XX DR WPI: 1989-312223/43.

XX DR P-PSDB: P91007.

XX PT Beta-mannase gene deoxyribonucleic acid

XX PS contg. recombinant plasmid

XX Claim 1; fig. 3; 10pp; Japanese.

CC This contains one restriction site each of XbaI, PstI, MclI, BglII, NcoI and NdeI. *E. coli* transformants produce beta-mannase efficiently. It is obtd. from *Bacillus* sp. FERM P-8856, -8857 and -8858.

XX Sequence 1461 BP; 475 A; 257 C; 340 G; 389 T; 0 other;

Query Match
Best Local Similarity 55.0%; Pred. No. 1e-21;
Matches 259; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

QY 910 tctgatgaggaactctccgacaactctttatgatttgaaggtgagatgcaagatg 969
DB 991 tatgtcaatcaatccacgacactgtctctatgatttgaataacacgctacg 1050
QY 970 actggaagatgagctt--gagcgagagcttggtgctgacagatggtcttaagga 1026
DB 1051 tccgggtggaatttcagcagcagagacatgaccttcgaatgagatgctgcaatg 1110
QY 1027 agtcaatctttaaagcggatattcaattgtcgtcaaatccacataacttaactg 1086
DB 1111 acctaatcgttgaagcagatgctgtcggcaataatagctacat--ttgcataaa 1167
QY 1087 atccaataacgctcttaccagcagaatgtagatcaagctactgttaacatgcaat 1146
DB 1168 acagtgaaatcgaaactctgtatcccaaaacctgaataatgaatgacatctctg 1227
QY 1147 tggggaagctgtgtaatgaaatgactgcgcgtcttattgtgaaacagagacatggtat 1206
DB 1228 tggggaatgtagaagtgatgacagcagaggtttcgtcaaaacagggagtgcttg 1287
QY 1207 acatgtaactctggaagcttctgacgattaaaggtctatctgcaacagcgtactct 1266
DB 1288 agatgaaatcaggtgaatttctgcgttgcggcgaagcaacgcgacatctcat 1347
QY 1267 gattatcaatctccaaactctctcaagtaaggaattggaatgcttcaatca 1326
DB 1348 gatttgacgaagtaagtaactctgcgatgctgcgagagtagtgtagatataagaa 1407
QY 1327 gcgagatgagtagtgacaaacatcgattatataatgattgattgta 1377
DB 1408 ccagcaaatagcaacggaagcagcgatctacttagatcatgtgacgta 1458

RESULT 13
245344 ID 245344 standard; DNA; 397 BP.

XX 245344;

DT 27-MAR-2000 (first entry)

DE DNA encoding a partial Bacillus mannanase enzyme.

XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture; ss.

OS Bacillus clausii.

XX Location/Qualifiers

XX Key 1..396

XX CDS

XX /tag= "mannanase"
XX /product= "partial sequence"

XX WO964619-A2.

XX 16-DEC-1999.

XX

PF 10-JUN-1999; 99WC-DK00314.
XX 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105870.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.

(NOVO) NOVO-NORDISK AS.

XX Kaupinen MS, Schnelein M, Schorr K, Andersen LN, Bjornvad ME;

XX WPI; 2000-105891/09.

DR P-PSDB; Y54131.

PT New mannanases for treatment of textiles, plant material and coffee
extract, and in cleaning compositions

PS Disclosure; Page 226-227; 242pp; English.

XX The present sequence encodes a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannanases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannan, galactomannan,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.

SO Sequence 397 BP; 113 A; 69 C; 116 G; 99 T; 0 other;

Query Match 7.6%; Score 104.4; DB 21; Length 397;
Best Local Similarity 57.3%; Pred. No. 1.6e-20;
Matches 189; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 4 gcaatccgatttatgtaagcgttacactctatagatgcgaatggaaccatt 63
DB 67 ggcacaacggatttcaagtaacagcgaccaggttctgtagagagggcaatcgat 126
QY 64 gtaatgagaggaattacaatggcgcagcatgtatataagaccggcaactctcaatt 123
DB 127 gtatgctgtgagtcataatcagacatcatatgttcaacaagacattgatacgaata 186
QY 124 gaaggatttgcacaatccggtgtaatacgtccgcatgtgtatctatgtaggagaca 183
DB 187 ccagctattgcagcagcgtcgcttaatacgttgagagatctgtttatcgatggcaaca 246
QY 184 tggacaaaagatgacatccatcagtaagaacattatctcttagcggaaataatcat 243
DB 247 tgggagcagagatcacgttagcggagtgtaagagtgcttgacgtaacgaagaggaagc 306
QY 244 ttgtgtctgttttgaagttatgtagtaccggttatgattccattcttgcgcaat 303

| Matches | 82: Conservative | 0: Mismatches | 67: Indels | 0: Gaps |
|-----------|-------------------------------------------------------------------------|---------------------------------------------------------|------------|---------|
| QY | 806 attgagctcgggaataaactcttaacagcttgggggaatacatagtcgaatgtrccataggt | 865 | | |
| DB | 376 atggtcgtggaattaccctctcttagtgcgtatcatgaattatttcccatatgcat | 435 | | |
| QY | 866 taaggaacttcgcgattgaagcacgcgttttaacagctggaggaatctgatatggaggaact | 925 | | |
| DB | 436 taatgcgtcttattattattatcatcgcgtgcatatttaagtcggaatgacugtggtcgaact | 495 | | |
| QY | 926 ctccgacacactcttatgatttgaaggt | 954 | | |
| DB | 496 atcaagcacacaggaatttgatttaaaagt | 524 | | |
| RESULT 15 | | | | |
| XX | W74586/c | | | |
| XX | W74586 standard; DNA; 4826 BP. | | | |
| XX | V74586; | | | |
| XX | 16-MAR-1999 (first entry) | | | |
| DE | Staphylococcus aureus contig SEQ ID #275. | | | |
| XX | Computer readable medium; vaccine; S.aureus infection; immunodetection; | | | |
| KW | cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy; | | | |
| KW | skin infection; surgical wound infection; scalded skin syndrome; | | | |
| KW | toxic shock syndrome; ds. | | | |
| XX | Staphylococcus aureus. | | | |
| OS | | | | |
| XX | Key | Location/Qualifiers | | |
| FH | misc_feature | 1081..1140 | | |
| FT | | /*tag= a | | |
| FT | | /note= "these bases represent a line of missing text in | | |
| FT | | the sequence listing in the specification. They | | |
| FT | | are included to maintain the nucleotide numbering | | |
| FT | | given in the specification for this DNA sequence" | | |
| FT | misc_feature | 2881..2940 | | |
| FT | | /*tag= b | | |
| FT | | /note= "these bases represent a line of missing text in | | |
| FT | | the sequence listing in the specification. They | | |
| FT | | are included to maintain the nucleotide numbering | | |
| FT | | given in the specification for this DNA sequence" | | |
| FT | misc_feature | 4681..4740 | | |
| FT | | /*tag= c | | |
| FT | | /note= "these bases represent a line of missing text in | | |
| FT | | the sequence listing in the specification. They | | |
| FT | | are included to maintain the nucleotide numbering | | |
| FT | | given in the specification for this DNA sequence" | | |
| XX | EP786519-A2. | | | |
| XX | 30-JUL-1997. | | | |
| XX | 07-JAN-1997; | 97EP-0100117. | | |
| XX | 05-JAN-1996; | 96US-0009861. | | |
| XX | | | | |
| PA | (HMAN-) HUMAN GENOME SCI INC. | | | |
| XX | Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA; | | | |
| XX | Rosen CA; | | | |
| XX | WPI: 1997-374922/35. | | | |
| XX | | | | |
| XX | Polynucleotide(s) and proteins derived from Staphylococcus aureus | | | |
| XX | stored on computer readable medium and used in the production of | | | |
| XX | anti-S.aureus vaccines | | | |
| XX | Claim 1: Page 1089-1092; 3271bp; English. | | | |

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the *S. aureus* DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against *S. aureus* infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the *S. aureus* DNA sequences contained on the
 CC computer readable medium.

CC
 XX
 SQ Sequence 4826 BP; 1645 A; 707 C; 893 G; 1400 T; 181 other;

Query Match 3.0%; Score 41.8; DB 18; Length 4826;

Best Local Similarity 55.0%; Pred. No. 0.045; Mismatches 67; Indels 0; Gaps 0;

Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 806 attggcctggaataacacctacagcttggggaataacatagtgatgccatagt 865
 DB 348 ATGGTCTTGGAATATCCACTCTTAGGTGATTCATGAATTTATCCCATATGAT 289
 QY 866 taagagaacctcgagattaagaccgctttacaggctgagagatctgatgagaact 925
 DB 288 TATGCGTCTTATATTTATCGCTGTTATTAGTGAATGACTGTTGCAACT 229
 QY 926 ctccgaacctcttatgatttgaagt 954
 DB 228 ATCAAGCAACAGGATTTGATTTAAAGT 200

Search completed: December 20, 2000, 09:59:18
 Job time: 30632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:56:44 ; Search time 75.74 Seconds
(without alignments) updates/sec
2735.645 Million cell

Title: US-09-339-159-1_COPY_91_1470

Perfect score: 1380
Sequence: 1 aatgcaatccgatttta.....ttgataatgattgtagaa 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/lna/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/1/lna/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/1/lna/6_COMB.seq:*
- 6: /cgn2_6/ptodata/1/lna/PCRTUS_COMB.seq:*
- 7: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| C 1 | 48.6 | 3.5 | 7218 | 1 US-08-232-463-14 | Sequence 14, Appl |
| 2 | 33.4 | 2.4 | 1365 | 4 US-08-870-827-4 | Sequence 4, Appl |
| 3 | 33.4 | 2.4 | 2408 | 4 US-08-870-827-5 | Sequence 5, Appl |
| 4 | 33 | 2.4 | 782 | 1 US-08-261-825-1 | Sequence 1, Appl |
| 5 | 33 | 2.4 | 782 | 3 US-08-719-124-1 | Sequence 1, Appl |
| 6 | 33 | 2.4 | 782 | 6 PCT-US95-07748-1 | Sequence 1, Appl |
| 7 | 33 | 2.4 | 782 | 6 PCT-US95-07748-1 | Sequence 1, Appl |
| 8 | 33 | 2.4 | 1800 | 5 US-07-745-206A-12 | Sequence 12, Appl |
| C 9 | 32.4 | 2.3 | 5467 | 3 US-08-311-363-12 | Sequence 12, Appl |
| 10 | 32.4 | 2.3 | 3467 | 3 US-08-566-398-23 | Sequence 23, Appl |
| 11 | 32 | 2.3 | 4359 | 6 PCT-US91-08525-21 | Sequence 21, Appl |
| 12 | 32 | 2.3 | 4365 | 6 PCT-US91-08525-25 | Sequence 25, Appl |
| 13 | 32 | 2.3 | 4365 | 6 PCT-US93-04384-1 | Sequence 1, Appl |
| 14 | 32 | 2.3 | 4365 | 6 PCT-US93-04384-7 | Sequence 7, Appl |
| 15 | 32 | 2.3 | 4365 | 6 PCT-US93-04384-15 | Sequence 15, Appl |
| 16 | 32 | 2.3 | 4365 | 6 PCT-US93-04384-15 | Sequence 15, Appl |
| 17 | 32 | 2.3 | 6090 | 3 US-08-566-398-35 | Sequence 35, Appl |
| 18 | 32 | 2.3 | 6144 | 7 US-08-566-398-32 | Sequence 32, Appl |
| C 19 | 31.8 | 2.3 | 7366 | 7 5169760-3 | Patent No. 5169760 |
| 20 | 31.8 | 2.3 | 7852 | 5 US-08-836-022A-2 | Sequence 2, Appl |
| C 21 | 31.8 | 2.3 | 7897 | 5 US-08-836-022A-1 | Sequence 1, Appl |
| 22 | 31.8 | 2.3 | 8509 | 2 US-08-462-014-1 | Sequence 1, Appl |
| C 23 | 31.8 | 2.3 | 10398 | 3 US-08-331-384-1 | Sequence 1, Appl |
| C 24 | 31.8 | 2.3 | 10398 | 3 US-08-708-188-1 | Sequence 1, Appl |
| C 25 | 31.8 | 2.3 | 10398 | 3 US-08-836-087-1 | Sequence 1, Appl |
| C 26 | 31.8 | 2.3 | 19182 | 4 US-08-850-880-11 | Sequence 11, Appl |

| | | | | | |
|------|------|-----|-------|---------------------|-------------------|
| C 27 | 31.8 | 2.3 | 19182 | 4 US-08-944-916-11 | Sequence 11, Appl |
| C 28 | 31.6 | 2.3 | 40328 | 5 US-08-742-185-102 | Sequence 102, App |
| C 29 | 31.6 | 2.3 | 43795 | 5 US-08-742-185-101 | Sequence 101, App |
| 30 | 31.4 | 2.3 | 8855 | 3 US-08-542-003-1 | Sequence 1, Appl |
| 31 | 31.4 | 2.3 | 8855 | 3 US-08-322-760A-1 | Sequence 1, Appl |
| 32 | 31 | 2.2 | 1453 | 4 US-08-252-995D-1 | Sequence 1, Appl |
| 33 | 31 | 2.2 | 1453 | 4 US-08-834-108-1 | Sequence 1, Appl |
| 34 | 31 | 2.2 | 1600 | 4 US-08-252-995D-5 | Sequence 5, Appl |
| 35 | 31 | 2.2 | 1600 | 4 US-08-834-108-5 | Sequence 5, Appl |
| 36 | 31 | 2.2 | 2882 | 1 US-08-424-567-1 | Sequence 1, Appl |
| 37 | 31 | 2.2 | 2882 | 3 US-08-711-928-1 | Sequence 1, Appl |
| 38 | 31 | 2.2 | 3447 | 1 US-08-252-995D-3 | Sequence 3, Appl |
| 39 | 31 | 2.2 | 3447 | 4 US-08-834-108-3 | Sequence 3, Appl |
| C 40 | 30.6 | 2.2 | 18994 | 2 US-08-459-586-4 | Sequence 4, Appl |
| C 41 | 30.6 | 2.2 | 18994 | 3 US-08-282-696-4 | Sequence 4, Appl |
| C 42 | 30.2 | 2.2 | 5718 | 5 US-08-714-918-48 | Sequence 48, Appl |
| C 43 | 30 | 2.2 | 2675 | 1 US-07-749-001-4 | Sequence 4, Appl |
| C 44 | 30 | 2.2 | 2675 | 1 US-08-154-198-4 | Sequence 4, Appl |
| C 45 | 30 | 2.2 | 2675 | 1 US-08-463-335-4 | Sequence 4, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: PARKER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)833-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-F1s
US-08-232-463-14

Query Match 3.5%; Score 48.6; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 0.00011;
Matches 15; Conservative 216; Mismatches 160; Indels 0; Gaps 0;

QY 657 agcatgtggaatttgacacgcgtcatcaaatgtgacgtgacgaacacgttat 716
DB 1455 AGAGATGAGAGATTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1396
QY 717 gaactatctgaacaagaagattggtgtgtgtgtgtgtgtgtgtgtgtgtgt 776
DB 1395 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1336
QY 777 agaattggagatttgaccttcgaatgattggtgtgtgtgtgtgtgtgtgt 836
DB 1335 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1276
QY 837 aatatacagtagatggtcctcatatgttttaagaagaactcgagattagccgtttt 896
DB 1275 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1216
QY 897 tacagtgagagatctgattgaggaactctccgacaactcttaattgtaggtag 956
DB 1215 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1156
QY 957 tatgcaagagatgagtagagtagctgtgagcgaggtcctgtgtgtgtgtgtgt 1016
DB 1155 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1096
QY 1017 ttctaagaagaactcattcttaagaacgagat 1047
DB 1095 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1065

RESULT 2

US-08-870-827-4
; Sequence 4, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1365 base pairs
; TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
US-08-870-827-4

Query Match 2.4%; Score 33.4; DB 4; Length 1365;
Best Local Similarity 55.7%; Pred. No. 1.7;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 172 gatgggagacaattgacaaagatgacatcatcagtaagaacctatctttagcg 231
DB 655 GATGAAGGCAAGTGGATTTCAGGTGATACCATCTTAAAGAGACCTCACTATGTTGAA 714
QY 232 gaagataatcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 286
DB 715 GATAAGGCCATAAATATCTGTCTTTGAAGCATCTGAAACACAGATGTT 769

RESULT 3

US-08-870-827-5
; Sequence 5, Application US/08870827
; Patent No. 5962297
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI et al.
; TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
; TITLE OF INVENTION: ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,827
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 170,630/1996
; FILING DATE: 10-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORGANISM: Bacillus sp.
; INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..360
; IDENTIFICATION METHOD: E
; NAME/KEY: signal peptide
; LOCATION: 361..456
; IDENTIFICATION METHOD: S
; NAME/KEY: mat peptide
; LOCATION: 457..1821
; IDENTIFICATION METHOD: S

NAME/KEY: 3'UTR
LOCATION: 1822..2408
IDENTIFICATION METHOD: E
US-08-870-827-5

Query Match 2.4%; Score 33.4; DB 4; Length 2408;
Best Local Similarity 55.7%; Pred. No. 2.2;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 172 gatggggagcaatgagcaaaagatgacatccatagtaagaaccttctcttagcg 231
DB 1111 GATGAAGGCAAGGATTCAGGTGATTAACACTTTAAGACACCTCATGTTGA 1170
QY 232 gaagataatcattgctgctcttctgaagttcatgatcgctcgtaattgatt 286
DB 1171 GATAAGGCCATTAATATCTTGTCTTGAAGCAATACGACACAGATGGTT 1225

RESULT 4

US-08-261-825-1
Sequence 1, Application US/08261825
Patent No. 5558993
GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261.825
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-261-825-1

Query Match 2.4%; Score 33; DB 1; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgccaaaatcttctcagtaagggaattgagttccatccagcgagta 1334
DB 35 AAATTACCAAAATCTTAATCAACAAATGCTTAATTGCGAACACCAAAATAGCTATTGT 94
QY 1335 tagtagtgacaacatcgattatattgataatgagttgata 1379
DB 95 TAGTCTGTCTTAATCTTATTAGTATTGCTAATAAGATGTATA 139

RESULT 5

US-08-719-124-1
Sequence 1, Application US/08719124
Patent No. 5854395
GENERAL INFORMATION:
APPLICANT: Champion, Cheryl I.
APPLICANT: Lovett, Michael A.
APPLICANT: Haake, David A.
APPLICANT: Miller, James N.
APPLICANT: Bianco, David R.
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719.124
FILING DATE: 24-SEP-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261.825
FILING DATE: 17-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: TUMARKIN, LISA A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD3516
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
US-08-719-124-1

Query Match 2.4%; Score 33; DB 3; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgccaaaatcttctcagtaagggaattgagttccatccagcgagta 1334
DB 35 AAATTACCAAAATCTTAATCAACAAATGCTTAATTGCGAACACCAAAATAGCTATTGT 94
QY 1335 tagtagtgacaacatcgattatattgataatgagttgata 1379

Db 95 TAGCTGTCTTAATCTATTAATGCAATAGAACTATA 139

RESULT 6
PCT-US95-07748-1
Sequence 1, Application PC/TUS9507748
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07748
FILING DATE: 16-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07419/013W01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
PCT-US95-07748-1

Query Match 2.4%; Score 33; DB 6; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgccaacaccttcctcaagtaaggaaattgagttcgaatcagcgagtga 1334
Db 35 AAATTAACCAAAATCTTAATCAACAAATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1335 tagtagtgacaacacatcgattatattgataatgattgtaga 1379
Db 95 TAGCTGTCTTAATCTATTAATGCAATAGAACTATA 139

RESULT 7
PCT-US95-07748A-1
Sequence 1, Application PC/TUS9507748A
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: CLONED Borrelia burgdorferi VIRULENCE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07748A
FILING DATE: 16-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, Ph.D., LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07419/013W01 (PD3516)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 782 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: Eppa
FEATURE:
NAME/KEY: CDS
LOCATION: 159..680
PCT-US95-07748A-1

Query Match 2.4%; Score 33; DB 6; Length 782;
Best Local Similarity 57.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1275 aaatgccaacaccttcctcaagtaaggaaattgagttcgaatcagcgagtga 1334
Db 35 AAATTAACCAAAATCTTAATCAACAAATGCTTAATTGCGAACCAAAATAGACTATTGT 94
QY 1335 tagtagtgacaacacatcgattatattgataatgattgtaga 1379
Db 95 TAGCTGTCTTAATCTATTAATGCAATAGAACTATA 139

RESULT 8
US-09-039-773A-1
Sequence 1, Application US/09039773A
Patent No. 6100388
GENERAL INFORMATION:
APPLICANT: Casas, Ivan
APPLICANT: Jonsson, Hans
APPLICANT: Mvlistam, Bo
APPLICANT: Roos, Stefan
TITLE OF INVENTION: Lactobacilli Harboring Aggregation and Mucin
TITLE OF INVENTION: Binding Genes As Vaccine Delivery Vehicles
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Standley & Gilcrest
STREET: 495 Metro Place South, Suite 210
CITY: Dublin
STATE: Ohio
COUNTRY: US
ZIP: 43017
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS Version 6.22
SOFTWARE: Microsoft Word Version 6.0

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/039,773A
: FILING DATE: 16-MAR-1998
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA: No. 6100388 applicable
: ATTORNEY/AGENT INFORMATION:
: NAME: Donald G. Nickey
: REGISTRATION NUMBER: 29,092
: REFERENCE/DOCKET NUMBER: 1229-005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (614) 792-5555
: TELEFAX: (614) 792-5536
: TELETYPE: No. 6100388 applicable
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1800 base pairs
: TYPE: Nucleic acid
: STRANDEDNESS: Double
: TOPOLOGY: Circular
: MOLECULE TYPE: Genomic DNA
: DESCRIPTION: Genomic DNA sequence and deduced amino
: DESCRIPTION: acid sequence of bacterial aggregation
: HYPOTHEICAL: No
: ANTI-SENSE: Yes
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: ORGANISM: Lactobacillus reuteri sp
: STRAIN: 1063
: CELL TYPE: Unicellular organism
:
: US-09-039-773A-1

```

```

Query Match      2.4%; Score 33; DB 5; Length 1800;
Best Local Similarity 53.5%; Pred. No. 2.5;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 507 acatcagattcatatgataagagaagtttatactgacccctcaagaataacat 566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1293 AGAAGCATTTTAAAGGCTCAAGATCGCGCTTAAATGATATCGATTAATCGCGCA 1352

QY 567 gtctcgattcatatgataagagaagtttatactgacccctcaagaataacat 626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1353 GGATTCACATCATCTGTTATGTAAGAACCGCTGAAGAGCTATTAGAACTCATATGCAAC 1412

QY 627 tgaccgaagt 635
    ||| ||| |||
DB 1413 TGACCTAGT 1421

RESULT 9
US-07-745-206A-12/c
: Sequence 12, Application us/07745206A
: Patent No. 5429921
: GENERAL INFORMATION:
: APPLICANT: Harpold, Michael
: APPLICANT: Ellis, Steven
: APPLICANT: Williams, Mark
: APPLICANT: McCue, Ann
: APPLICANT: Feldman, Daniel
: TITLE OF INVENTION: Human Calcium Channel Compositions and
: TITLE OF INVENTION: Methods
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fitch, Even, Tabin & Flannery
: STREET: 135 S. LaSalle
: CITY: Chicago
: STATE: Illinois
: COUNTRY: U.S.A.
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

:
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/745,206A
: FILING DATE: 19910815
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Feder, Scott B
: REFERENCE/DOCKET NUMBER: 51504
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-372-7842
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5467 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(144..3164, 3168..3245, 3249..3386, 3390
: LOCATION: ..3392, 3396..3488, 3495..3539, 3543..3581, 3585
: LOCATION: ..3587, 3591..3626, 3630..3689, 3693..3737, 3744
: LOCATION: ..3746, 3750..4823, 4827..4841, 4845..5006, 5010
: LOCATION: ..5096, 5100..5306, 5310..5366, 5370..5465)
:
: US-07-745-206A-12

```

```

Query Match      2.3%; Score 32.4; DB 1; Length 5467;
Best Local Similarity 54.1%; Pred. No. 6.3;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 1185 ttgtaaacagagacatggtatcatatgactctggaagctttgcccgaattaacggttc 1244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4884 TGGTAAGAGAGAGTCAAGCTACCTCGTGTGCTGATACATTAAACCTGGAACCTTC 4825

QY 1245 atctggaacacagctatctcagattatcaaatgctcaaatcttctcaagtaaggga 1304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4824 ATCTGGGCAATGCAATCTCTTACTAAATATCATGTAATCTTCCCAACACAGTGA 4765

QY 1305 aa 1306
DB 4764 CA 4763

RESULT 10
US-08-311-363-12/c
: Sequence 12, Application us/08311363
: Patent No. 5876958
: GENERAL INFORMATION:
: APPLICANT: Harpold, Michael
: APPLICANT: Ellis, Steven
: APPLICANT: Williams, Mark
: APPLICANT: Feldman, Daniel
: APPLICANT: McCue, Ann
: APPLICANT: Brenner, Robert
: TITLE OF INVENTION: Human Calcium Channel Compositions and
: TITLE OF INVENTION: Methods
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-2926
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/311,363

```

FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-51506
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)238-0999
TELEFAX: (619)238-0062
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 5467 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(144..3164, 3168..3245, 3249..3386, 3390..3392, 3396..3488, 3495..3539, 3543..3581, 3585..3587, 3591..3626, 3630..3689, 3693..3737, 3744..3746, 3750..4823, 4827..4841, 4845..5006, 5010..5096, 5100..5306, 5310..5366, 5370..5465)
US-08-311-363-12

Query Match 2.3%; Score 32.4; DB 3; Length 5467;
Best Local Similarity 54.1%; Pred. No. 6.3;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 1185 tctgaagaagacatgtatatacatgtacatctggaagcttggcgaattaaagcttc 1244
DB 4884 tctgaagaagacatgtatatacatgtacatctggaagcttggcgaattaaagcttc 4825
QY 1245 atctggaacacgctatctctgattatcaaatgtccaaatcttctcaagtaaggga 1304
DB 4824 atctggaacacgctatctctgattatcaaatgtccaaatcttctcaagtaaggga 4765
QY 1305 aa 1306
DB 4764 CA 4763

RESULT 11
US-08-566-398-23
Sequence 23 Application US/08566398
Patent No. 5658373
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: Gettly, Russell
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIONS
TITLE OF INVENTION: PERIONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/566,398
FILING DATE: 01-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2880
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 4359 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-566-398-23

Query Match 2.3%; Score 32; DB 3; Length 4359;
Best Local Similarity 65.3%; Pred. No. 7.5;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 520 gattatgaagaagattttaaagctgaccccaacgaataatgtttgattcat 579
DB 2587 gattatgaagaagattttaaagctgaccccaacgaataatgtttgattcat 2646
QY 580 atgtatgaat 591
DB 2647 GTGTCTGCATGT 2658

RESULT 12
PCT-US91-08525-21
Sequence 21, Application PC/TUS9108525
GENERAL INFORMATION:
APPLICANT: SmithKline Beecham, Corporation
TITLE OF INVENTION: Recombinant Feline Coronavirus S
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08525
FILING DATE: 19911114
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: King, William T.
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: SBC 14532B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA

Oy 520 gattatggaagaagaagttttaaigtgcaccctcaacgaatacatalgttcgatccat 579
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2593 GATTGTGCAGATACGTTTGTGAATGTAACCTAGATGTAAACAATGTGTTAACACAAATAT 2655

00-339-159-1 copy 91 1470.rnj

2653 GTGTCGTCATGT 2664

Search completed: December 20, 2000, 09:57:17
Job time: 30824 sec

```

RESULT 15
PCT-US93-04384-7
Sequence 7, Application PC/TUS9304384
GENERAL INFORMATION:
APPLICANT: Miller, Timothy J.
APPLICANT: Klepfer, Sharon
APPLICANT: Reed, Albert Paul
APPLICANT: Jones, Elaine V.
TITLE OF INVENTION: Compositions and Methods for Vaccination
TITLE OF INVENTION: Against Coronaviruses
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation - Corporate
ADDRESSEE: Patents
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384
FILING DATE: 19930507
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SBC H85009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US93-04384-7

```

| | | | | |
|-----------------------|--------------|----------------|----------------|--------------|
| Query Match | 2.38; | Score 32; | DB 6; | Length 4365; |
| Best Local Similarity | 65.3%; | Pred. No. 7.5; | | |
| Matches 47; | Conservative | 0; | Mismatches 25; | Indels 0; |
| | | | | Gaps 0; |

[illegible]

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 01:58:28 ; Search time 1967.85 Seconds
(without alignments)
4335.835 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_1470
Perfect score: 1380
Sequence: 1 aatgcaaatcgcgatttta.....ttgataatgtgattgtagaa 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189884 seqs, 3091403243 residues
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est4a:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_esth1:*
64: em_esth2:*
65: em_esth3:*
66: em_esth4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estp11:*
71: em_estp12:*
72: em_estp13:*
73: em_estp14:*
74: em_estp15:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*


```

Db 954 KTAGAGGCKKKKKKDDKRDADAKATGKXTWRTTPTAKRRRRRRRRAGRGDRKD 1013
Oy 972 tgaagtagcttgagcgaagtccttgctgctgcacagatgcttctaaagaaagta 1031
Db 1014 KKKRRARWRGKKKKKTDDGKRWGRRAKADAKADADADAADAAAWAKAKADK 1073
Oy 1032 ttctttaaagcgaat 1049
Db 1074 DKAKKADADDRKDKAK 1091

RESULT 2
LOCUS CNS00396/c 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPI1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921 GI:4941778
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI1-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="RPI1-98"
/clone="BACR08K10"
/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN
Query Match 3.3% Score 46.2; DB 121; Length 1101;
Best Local Similarity 14.5% Pred. No. 0.05;
Matches 65; Conservative 215; Mismatches 164; Indels 3; Gaps 1;

Oy 683 atacaatggtgacgtcgaagaacgatatgagctatctgaacaagagagtg 742
Db 1026 WMMATATKATJTATWMTATWTRADWAGRGAGKRRDADADAGRRDGRKRD 967
Oy 743 ggtgttggtgctgctgcaagaagagagagagagagagagagagagagagag 802
Db 966 RKDGDGDKGKKKAAATATWMDWMDKDWKMGADKRAADDGDDGDDGK 907
Oy 803 atgattggtcgaataacattacagcttgaggaaatacagatgagtgccatag 862
Db 906 GKADDDDFDGRKDDDKKMDMDKAKGTWGDATWAAATWMMGMAADADMDTDAAD 847

```

```

Oy 863 gtttaaggaacttca---gattaagcacgcttttacaggtgagagattgag 919
Db 846 DMMADDDWAMWKKDDAMWAGARTADRDWDRAGKGGAKRRDRADKRAAD 787
Oy 920 gaacttcacgaactcttatgatttgagagtagatgaagagagagagagag 979
Db 786 DDAATWTTWTTTTRDTRDMMKKTWTRMAADRWDDDDDRRAGTAGKMTKR 727
Oy 980 gcttgagcgaagtccttgctgacagagtgctcttaagaagagagagagag 1039
Db 726 RWRKRDTRWDADADTDARDDRDRRGDAGAGKGTGRKRRDRATWRTDAMWDA 667
Oy 1040 aagcgatattcaattgcgcgaatcacacattacttcatatttcaataagc 1099
Db 666 AAMTTTDDTDWDRDRRRKGAARRRRRTTAAADMTWAMDAKMDKTRADRWDR 607
Oy 1100 cttaacagcagatagtagtacag 1126
Db 606 AADTTDARKADRWAKARARARDR 580

RESULT 3
LOCUS CNS00LT2 1101 bp DNA GSS 14-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR48P19 of RPI1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL078714 GI:5102004
VERSION AL078714.1 GI:5102004
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI1-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="RPI1-98"
/clone="BACR48P19"
/note="end : TET3"

BASE COUNT 469 a 6 c 69 g 151 t 406 others

ORIGIN
Query Match 3.1% Score 43.4; DB 121; Length 1101;
Best Local Similarity 14.9% Pred. No. 0.3;
Matches 48; Conservative 140; Mismatches 135; Indels 0; Gaps 0;

Oy 225 tttagcgaataatacatcttggtgctgcttgagatcatgctaccggtatga 284

```

```

Db 699 TTKKKKKKAAAADKWDAAKWDAAKATKTKKKDKKAAKAAADKRRKKGGKGGKGG 758
Qy 285 ttccattgcttcgctcaatcgctcgttgcattatgattgaaatgagaagtgcttaac 344
Db 759 KKGKKKKKKKKKKGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 818
Qy 345 tggaaaggaagatacgcctcattatattggaatgaaatgattgcttcgctgggaag 404
Db 819 KKAARTRTKTKWDAAAAAAAKKTRDKGKKKKTKTKTKKKKKKKKKKKGGGGKAAK 878
Qy 405 ggaattgctgctgacgaggtataaacaagcaatcccgatgctgaacgcgctcaaa 464
Db 879 KKKGGTKKKKKKKGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 938
Qy 465 ccatccttgatgtagatgctgctgaggtgaggaacattcccaatcgattcatgata 524
Db 939 KGGGKTKTKTKTKKKKKKKKADAAKGGTKKRAAADAADWTDAATKKKKKKKKDKKKT 998
Qy 525 tggaaagagaattttaaactgctg 547
Db 999 KTKKKDAAAATKTKKKKKKKTK 1021

```

```

RESULT 4
CNS0182P 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108811.1 GI:5629115
VERSION AL108811.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
GENOSCOPE.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

```

```

FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_1lb="DrosBAC"
/clone="BACN37D10"
/note="end : SP6"
BASE COUNT 274 a 268 c 128 g 73 t 358 others
ORIGIN

```

```

Query Match 3.1%; Score 43.4; DB 121; Length 1101;
Best Local Similarity 16.1%; Pred. No. 0.3;
Matches 66; Conservative 174; Mismatches 169; Indels 0; Gaps 0;
Qy 578 atatgtaataatgacgagtgtaatgacgcaatgctgactaataatgacgagtc 637
Db 692 ADARAAAAAAGAAAAAAGAGGKRWKKBGDKGKATKTAAMAARKNMGRTATA 751
Qy 638 ttaataaagactcgcattgattgattgattgacacgcctcacaatggtgacg 697

```

```

Db 752 WWTWATWADWTWKAATDADARAAGRRKRDARARARARRARRARRARRAG 811
Qy 698 tgcataagcaacgattatgactatctgaacaagaagagtggtgctgacggtg 757
Db 812 AARAAARAAADDDRWDAWMAAAAAAATTTWRBMDMDMDTTRMDDTTAAMD 871
Qy 758 catgaaaggaagcgcccgcaatgagatttagaccttcgaatgattgctggaa 817
Db 872 DARARARRRRRRRRARRARRAADDTDKRWADATDKDTTKWTDDDDWDKAK 931
Qy 818 ataaccatgactgctgggaataacatagatgctccatagtttaagaagaactt 877
Db 932 KDRWMAARADGAKWKRDRARDDMAATARDGDKWGRKRGKDKKDKKDKK 991
Qy 878 cgaattagacacgcttttcaagtgagatgctgataagtgaggaactccgcaactc 937
Db 992 DDDWKTWTBDMWMTTKMDWMDGGRGRTWRKRWGAWRAADARDDTGKDTRT 1051
Qy 938 tttaatttgaagtagtatgcaagatgactggaagtagctgag 986
Db 1052 ADKRRRTDTYKRDGDWKRDKRDKRRDKKDDTKKADATDDDDARD 1100

```

```

RESULT 5
CNS01412 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN1119 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL104180.1 GI:5615791
VERSION AL104180.1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
GENOSCOPE.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

```

```

FEATURES
source 1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_1lb="DrosBAC"
/clone="BACN1119"
/note="end : SP6"
BASE COUNT 380 a 159 c 215 g 262 t 85 others
ORIGIN

```

```

Query Match 3.1%; Score 42.2; DB 121; Length 1101;
Best Local Similarity 40.8%; Pred. No. 0.63;
Matches 125; Conservative 30; Mismatches 150; Indels 1; Gaps 1;
Qy 985 agcgagagtccttgagtgacagagtgctctcaagaagagtagtctttaaagc 1044
Db 722 AGCCAGGCTTWWKGGTTCAAGWACAAAAACCCCAAAATGAGGTGGGAAATGATGAACGG 791
Qy 1045 gataatcatgtcgtcaaatcaacattactatcatgttatcaaatagctctta 1104

```

| | | | |
|----|------|------------------------------------------------------------------|------|
| Db | 782 | GASACGCTGTAG-CCTTAAAGCTCGMTTTTKGCATCCGAKCYACAAMAAAATTTCATGAG | 840 |
| Oy | 1105 | cagacgaatagttagatacaagaactactcgtttaacaatgysaaatttgystaat | 1164 |
| Db | 841 | CTGCAGCAACMAACMAAAGWAPCAACCAACCATGWATAAGTAGTGAGNAGAAAGAATGCANW | 900 |
| Oy | 1165 | ggaaatgactgcgcgtcttttctfgaaaacacgsgaatggttatcatcgttacttccttgsaac | 1224 |
| Db | 901 | TWAATTTTTGCMNYTYTRGTCTGAAMAATWTMTIDGTTTGTCGASCACCAACAAAGSNC | 960 |
| Oy | 1225 | tcttgcggaattaacggttcatctgysacaacgtactctagattattcaaatgycaa | 1284 |
| Db | 961 | TGGGMTTWANAACRKGAAACACGACTAGBAGASGCTTACTGMMCAGTCATTAGKCTCCMA | 1020 |
| Oy | 1285 | aactct 1290 | |
| Db | 1021 | AWAGTT 1026 | |

| | |
|------------|-------------------------------------------------------------------------------------|
| RESULT | 6 |
| AZ028169/c | |
| LOCUS | AZ028169 471 bp DNA GSS 25-PEB-2000 |
| DEFINITION | RPCIT-23-31608.TV RPCI-23 Mus musculus genomic clone RPCI-23-31608 DNA sequence. |
| ACCESSION | AZ028169 |
| VERSION | AZ028169.1 GI:7103553 |
| KEYWORDS | GSS. |
| SOURCE | house mouse. |
| ORGANISM | Mus musculus |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|--------------------|-------------------------------------------------------------------------|----------------------------------------------|---------|---------|
| 1 (bases 1 to 471) | Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatman, S., Akinret | | | |
| | B., Levins, M., McGinn, S., Tsagaye, G., Geer, K., Krol, M., de Jong, P | | | |
| | and and Fraser, C.M. | Mouse BAC End Sequences from Library RPCI-23 | | |
| | | Unpublished (1999) | | |
| | Other_GSSs: RPCI-23-31608..TJ | | | |

| TITLE | JOURNAL | COMMENT |
|-----------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------|
| Mouse BAC End Sequences from library RPCI-23 | Unpublished (1999) | |
| Other_GSSS: RPCI-23-31608.T1 | | |
| Contact: Shaying Zhao | | |
| Department of Eukaryotic Genomics | | |
| The Institute for Genomic Research | | |
| 9712 Medical Center Dr., Rockville, MD 20850, USA | | |
| Tel.: 301 838 0200 | | |
| Fax: 301 838 0208 | | |
| Email: szhaoe@ligr.org | | |
| Clones are derived from the mouse BAC library RPCI-23. For BAC | | |
| library availability, please contact Plietier de Jong | | |
| (plietierdejong.med.buifalo.edu). Clones may be purchased from | | |
| BACPAC Resources (http://bacpac.med.buifalo.edu/orderingframe.htm | | |
| or from Resea chr Genetics (info@resgen.com). BAC end page: | | |
| http://www.tlgr.org/rdb/bac_ends/mouse/bac_end_intro.html | | |
| Plate: 316 | row: 0 | column: 8 |
| Seq primer: T7 | | |
| Class: BAC ends. | | |

```

FEATURES
source
    Location/Qualifiers
      1..471
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPC1-23-31608"
        /clone_1lb="RPC1-23"
        /sex="Female"
        /lab_host="DH10B"
        /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
        brain genomic DNA was isolated and partially digested
        with a combination of EcoRI and EcoRI Methylase. Size
        selected DNA was cloned into the pBAC3.6 vector at the
        EcoRI sites. The ligation products were transformed into
        DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT
151 a      115 c      88 g      117 t
BRIGIN

```

[illegible]

| RESULT | 7 |
|------------|-------------------------------------------------------------------|
| LOCUS | C90049/c |
| DEFINITION | C90049 : 350 bp mRNA EST 20-APR-1998 |
| ACCESSION | C90049 Dictyostellium discoidium SS (H.Urushihara) Dictyostellium |
| VERSION | C90049 |
| KEYWORDS | Dictyostellium cDNA clone SSG140, mRNA sequence. |
| SOURCE | C90049.1 GI:3059669 |
| ORGANISM | EST. |
| | Dictyostellium discoidium. |
| | Dictyostellium discoidium |

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (basses 1 to 350)
Yoshino, R., Morito, T. and Tanaka, Y.
Developmental cDNA in *Dicystostellium discoidium*
Unpublished (1997)
Contact: Hiideko Urushihara

FEATURES Location/Qualifiers
SOURCE 1 350

```

- /organism="Dictyostelium discoideum"
- /strain="AX4"
- /db_xref="taxon:44689"
- /clone="SSG140"
- /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
- /dev_stage="slug"
BASE COUNT      174 a      49 c      31 g      96 t
ORIGIN

```

| | | | | |
|-----------------------|-----------------|-----------------|-----------|-------------|
| Query Match | 3.0%; | Score 41.2; | DB 36; | Length 350; |
| Best Local Similarity | 49.5%; | Pred. No. 0.84; | | |
| Matches 106; | Conservative 0; | Mismatches 108; | Indels 0; | Gaps 0; |

| | | | |
|----|------|------------------------------------------------------------------|------|
| Qy | 1047 | tattcaatttgcgtcaaatcccaacattactactacatgtaattcaaaatgaagctttaca | 1106 |
| | | | |
| Db | 307 | TATTATTATGTTTATTATTATTGTTTATTATTATTATTACTTTATATATAAATATTTTGG | 248 |
| | | | |
| Qy | 1107 | gcagaatagtggatgatacaagctactgtttaacatgcgaattvgggaaagtgttgtaatg | 1166 |
| | | | |
| Db | 247 | GTAATTTGTATGATGATAGATGGATTAATAAATAATTTAATTTGGTGTAGTGTATGTAATCA | 188 |
| | | | |
| Qy | 1167 | aatgactgcgagctcttaattgtcaaacacagacatggtttatcacatgtaactctggaagctt | 1226 |
| | | | |
| Db | 187 | ATTGATTATATCTTGATCTCTTTGGAAGGACCAATTAATTAATTCAAAGTTTGAATGTT | 128 |
| | | | |
| Qy | 1227 | tgtagcgaattaaaggtcatctcggaacaagcta | 1260 |
| | | | |
| Db | 127 | GCTTATATATTCATTTCATAGTGTAACCAATCTA | 94 |
| | | | |

| | |
|--------|---|
| RESULT | 8 |
|--------|---|

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BB388799.1 GI:9111642
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 303)
Kono, H., Aikawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci,
P., Endo, T., Fukuda, S., Fukuishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, Y., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyoosawa, H., Koijima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, N.,
Matuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owai, C., Saito, H., Sekai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamuta, T., Yamanaka, I.,
Yano, R., Yasuniishi, A., Yokota, T., Yoshida, K., Yoshiaki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)

TITLE
JOURNAL
COMMENT

Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel.: +81-298-36-9033
Fax: +81-298-36-9038
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagoka, S., Sasakawa,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
Source

1..303
location/Qualifiers
organism="Mus musculus"
db_xref="taxon:10090"
clone="C230058N01"
clone_1lb="RIKEN full-length enriched, 0 day neonate
cerebellum"
tissue-type="cerebellum"
dev_stage="0 day neonate"
lab_host="DH10B"
note="Site.1: SalI; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGACAGAGAGATCCAGAGCTCTTTTCTTTTTTTTNN 3'] . cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 479.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGACAGAGATCTCGAGGTAAATTAAATCAATCCCCCCCCTTC
3'] . cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from lambda-

| | |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ACCESSION | sapiens genomic clone Plate=3089 Col=24 Row=P, DNA sequence. |
| VERSION | A0621653 |
| KEYWORDS | A0621653.1 GI:5084045 |
| SOURCE | GSS. |
| ORGANISM | human. |
| TITLE | Homo sapiens |
| JOURNAL | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| MEDLINE | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. |
| REFERENCE | 1 (bases 1 to 520) |
| AUTHORS | Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,W.D. and Hood,L. |
| COMMENT | Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999) 99380589 Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel.: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones may be purchased from Research Genetics (info@resgen.com). BAC end Web Server: http://www.husc.washington.edu Plate: 3089 row: P column: 24 Seq primer: T7 Class: BAC ends High quality sequence stop: 520. Location/Qualifiers 1..520 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_plate="3089 Col=24 Row=P" /clone_id="CIT Approved Human Genomic Sperm Library D" /sex="male" /note="Organ: sperm; Vector: pBelBAC11; BAC Clones In E-Coli DH10B" |
| FEATURES | |
| source | |
| BASE COUNT | 146 a 100 c 129 g 143 t 2 others |
| ORIGIN | |
| Query Match | 2.8%; Score 39.2; DB 100; Length 520; |
| Best Local Similarity | 57.3%; Pred. No. 3.4; |
| Matches | 71; Conservative 0; Mismatches 53; Indels 0; Gaps 0; |
| OY | 223 tccttagcggaagaatcaattgattgttcgttctcctaagatcatgatgcacggatat 282 |
| Db | 211 TCCTTCGTGCATACAGACTTCTATTATTTTCTGTTAAAGTTCTCCAGGTTTT 270 |
| OY | 283 gatccattcgttcgcatactgcigtgttgatttttgttatgtaagtgaagtgacttta 342 |
| Db | 271 GTTTAAATTGGATGTTCAAGATTTCGTTCCGATTCAGATTGCACTTAAGAAGATGCAATTTA 330 |
| OY | 343 attg 346 |
| Db | 331 AATG 334 |
| RESULT_13 | |
| CNSO16E2/c | 1204 bp DNA GSS 26-JUL-1999 |
| LOCUS | Drosophila melanogaster genome survey sequence T7 end of BAC |
| DEFINITION | BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit |
| ACCESSION | fly), genomic survey sequence. |
| VERSION | Alt06628 |
| KEYWORDS | Alt06628.1 GI:5622852 |
| SOURCE | GSS. |
| ORGANISM | fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epnydroidea; Drosophilidae; Drosophila. |

| REFERENCE | AUTHORS | TITLE | JOURNAL | COMMENT |
|---------------------------------------------------------------------|-------------------------------------------------------------------------|---------------------------------------------------------------|-------------------------------------------------------------------|------------------|
| 1 (bases 1 to 1204) | Genoscope. | Direct Submission | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage | |
| BP 191 91006 EVR | cedex - FRANCE (E-mail : segretef@genoscope.cns.fr | | | |
| - Web : www.genoscope.cns.fr | | | | |
| - Delineation of this BAC-end sequence was carried out as part of a | | | | |
| collaboration with the European Drosophila Genome Project (EDGP) - | | | | |
| http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC | | | | |
| library (Dros BAC) was made by Alain Billard at CEPH (Centre | | | | |
| d'Etude du Polymorphisme Humain) with funding provided by a MRC | | | | |
| project grant. The DNA was prepared from embryos by Alain Bucheton | | | | |
| and Genevieve Payan. It has been constructed in the vector | | | | |
| pBelOBAC11. | | | | |
| FEATURES | source | Location/Qualifiers | | |
| | | 1..1204 | | |
| | | /organism="Drosophila melanogaster" | | |
| | | /plasmid="pBelOBAC11" | | |
| | | /db_xref="taxon:7227" | | |
| | | /clone_11b="DrosBAC" | | |
| | | /clone="BACN15A12" | | |
| | | /note="end : 17" | | |
| BASE COUNT | 298 a | 172 c | 106 g | 316 t 312 others |
| ORIGIN | | | | |
| Query Match | 2.8%; Score 39.2; DB 121; Length 1204; | | | |
| Best Local Similarity | 19.6%; Pred. No. 4.4; | | | |
| Matches | 65; Conservative 134; Mismatches 131; Indels 2; Gaps 1 | | | |
| OY | 1045 | gataatcaatgctgcgaattcaacattcaacattcctacatgattatccaatcgccttta | 1104 | |
| | | | | |
| Db | 1181 | KATAAKAWWKAAXKAKKAAAAAAMAAAAAADAAAAAKKRAAAAAAARADAA | 1122 | |
| OY | 1105 | cagcagaatag-taggatcaagctactcgtttaacaatgcaattgagggaagtgttgta | 1162 | |
| | | | | |
| Db | 1121 | AADAAAAAKKAKAKKAAAAAADAAMKKKAAAAAAMAAKKATKKDDDKDA | 1062 | |
| OY | 1163 | atggaatgactgcgcgtctttattgtgaanaacaggacatggtatatacattgctctctg | 1222 | |
| | | | | |
| Db | 1061 | DAKKKKKKAKKADADAAAKDKDKKAKKAAAMDADADAANKKKKKDKKKDDATD | 1002 | |
| OY | 1223 | gctttgctgcgattacaggttcattcattggaacaacgctactctagattatcaaatgctc | 1282 | |
| | | | | |
| Db | 1001 | KAACKKDKKAKKKKKDKKKAKKAKKAKKKDKKKDKKKKKDKKAKKAKKRAA | 942 | |
| OY | 1283 | aaaattcttctcaagtaagggaattgtgagttcagttccaatcagcagatgatagtg | 1342 | |
| | | | | |
| Db | 941 | AAAAAAKKDKKKDAKAKKKKKKDDAKKKKKKDDAMADADAKAAKKKAAKKK | 882 | |
| OY | 1343 | gacaaacatcgattatattgataatgattgatt | 1374 | |
| | | | | |
| Db | 881 | KDAADKRAKADAAKKKKDKRAKAAADKDAKK | 850 | |
| RESULT 14 | | | | |
| LOCUS | AZ138670 | 750 bp | DNA | GSS |
| DEFINITION | SP_0170_A2_C04_SP6E Strongylocentrotus purpuratus, purple sea | | | |
| | urchin, sperm genomic BAC library Strongylocentrotus purpuratus | | | |
| | genomic clone Plate-18 Col-8 Row-E, DNA sequence. | | | |
| ACCESSION | AZ138670 | | | |
| VERSION | AZ138670.1 | GI:8290573 | | |
| KEYWORDS | GSS. | | | |
| SOURCE | Strongylocentrotus purpuratus. | | | |
| ORGANISM | Strongylocentrotus purpuratus | | | |
| | Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; | | | |
| | Echinoidea; Euechinoidea; Echnacea; Echinoida; | | | |
| | Strongylocentrotidae; Strongylocentrotus. | | | |
| REFERENCE | 1 (bases 1 to 750) | | | |
| AUTHORS | Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Blondi, T.R., | | | |
| | Swartzell, I.S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray | | | |
| | , G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and | | | |

Wed Dec 20 15:35:13 2000

us-09-339-159-1_copy_91_1470.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 03:12:31 ; Search time 4256.96 Seconds

(without alignments)
923.387 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990

Sequence: 1 aatgcaaatccgagatttta.....gattaagaccggtttttaca 900

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_cm:*
20: em_cor:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl1:*
25: em_ro:*
26: em_sy:*
27: em_un:*
28: em_un:*
29: em_vl:*
30: gb_htg1:*
31: gb_htg2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba1:*
35: em_ba2:*
36: em_hum3:*
37: em_hum4:*
38: gb_pr4:*
39: gb_htg3:*
40: gb_htg4:*
41: gb_htg5:*
42: gb_htg6:*
43: gb_htg7:*

44: em_htg1:*
45: em_htg2:*
46: em_htg3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pr5:*
50: gb_htg8:*
51: gb_htg9:*
52: gb_htg10:*
53: gb_htg11:*
54: gb_htg12:*
55: gb_htg13:*
56: gb_htg14:*
57: gb_in3:*
58: gb_htg15:*
59: gb_htg16:*
60: gb_htg17:*
61: em_htg4:*
62: em_htg5:*
63: em_htg6:*
64: em_htg7:*
65: em_hum6:*
66: gb_htg18:*
67: gb_htg19:*
68: gb_htg20:*
69: gb_htg21:*
70: gb_htg22:*
71: gb_htg23:*
72: gb_vl1:*
73: gb_vl2:*
74: gb_ba3:*
75: em_htg8:*
76: em_htg9:*
77: em_htg10:*
78: em_htg11:*
79: em_htg12:*
80: em_htg13:*
81: em_htg14:*
82: em_htg15:*
83: em_htg16:*
84: em_htg17:*
85: em_htg18:*
86: em_htg19:*
87: em_htg20:*
88: em_htg21:*
89: em_htg22:*
90: em_htg23:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being predicted,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | DB ID | Description |
|------------|-------|---------|--------|-------|------------------------------|
| 1 | 422.4 | 46.9 | 1551 | 1 | AB007123 Bacillus |
| 2 | 239.2 | 26.6 | 1530 | 2 | D86329 Vibrio sp. |
| 3 | 134 | 14.9 | 4567 | 1 | AF163837 Caldicl |
| 4 | 131.8 | 14.6 | 2155 | 2 | CDOMANABD M36063 C.saccharol |
| 5 | 131.8 | 14.6 | 4977 | 2 | CDOMANA L01257 Caldocellum |
| 6 | 125 | 13.9 | 5284 | 2 | CS016508 U16308 Caldocellum |
| 7 | 91 | 10.1 | 25970 | 1 | SC2H4 AL031514 Streptomy |
| 8 | 85.8 | 9.5 | 1881 | 74 | STM8MANASE M92297 Streptomy |
| 9 | 81.6 | 9.1 | 837 | 74 | TFU6227 AJ006227 Thermomon |
| 10 | 52.4 | 5.8 | 14922 | 2 | SCF73 AL121746 Streptomy |
| 11 | 45.8 | 5.1 | 111071 | 53 | AC024591 NC024591 Homo sapi |
| 12 | 45.2 | 5.0 | 35049 | 33 | CELK0386 U55375 Caenorhabdi |

| | | | | | | |
|----|----|------|--------|--------|----------|-----------|
| 13 | 45 | 5.0 | 175300 | 40 | AC012459 | Homo sapi |
| C | 14 | 44 | 4.9 | 14235 | 73 | H2U67264 |
| C | 15 | 42.8 | 4.8 | 784867 | 38 | HS101859 |
| C | 16 | 42.8 | 4.8 | 251206 | 68 | AL354834 |
| C | 17 | 42.4 | 4.7 | 1063 | 72 | HA067261 |
| C | 18 | 41.8 | 4.6 | 214412 | 43 | AC018695 |
| C | 19 | 41.6 | 4.6 | 17660 | 48 | SC9723 |
| C | 20 | 41.6 | 4.6 | 38779 | 48 | SC931900 |
| C | 21 | 41.6 | 4.6 | 107642 | 12 | AF125313 |
| C | 22 | 41.6 | 4.6 | 165536 | 48 | SC931900 |
| C | 23 | 41.4 | 4.6 | 3734 | 8 | ATRNAHE1 |
| C | 24 | 41.4 | 4.6 | 81493 | 8 | AT81KBEN |
| C | 25 | 41.4 | 4.6 | 82646 | 7 | AB028611 |
| C | 26 | 40.8 | 4.5 | 1302 | 11 | AF090946 |
| C | 27 | 40.8 | 4.5 | 56804 | 49 | HS77919 |
| C | 28 | 40.2 | 4.5 | 2103 | 57 | TM049024 |
| C | 29 | 40.2 | 4.5 | 44101 | 60 | AC073860 |
| C | 30 | 40 | 4.4 | 14222 | 1 | AE004401 |
| C | 31 | 39.8 | 4.4 | 171766 | 41 | AC016048 |
| C | 32 | 39.6 | 4.4 | 2294 | 57 | YMC09206 |
| C | 33 | 39.4 | 4.4 | 11379 | 48 | MRAJ2479 |
| C | 34 | 39.4 | 4.4 | 187615 | 50 | AC020724 |
| C | 35 | 39.4 | 4.4 | 237523 | 2 | RFX04 |
| C | 36 | 39 | 4.3 | 2318 | 33 | AF064521 |
| C | 37 | 39 | 4.3 | 145011 | 31 | AC009885 |
| C | 38 | 39 | 4.3 | 147700 | 70 | AC002362 |
| C | 39 | 39 | 4.3 | 157591 | 43 | AC018775 |
| C | 40 | 39 | 4.3 | 164991 | 11 | AC011088 |
| C | 41 | 39 | 4.3 | 167482 | 70 | AP000795 |
| C | 42 | 39 | 4.3 | 169168 | 70 | AP002497 |
| C | 43 | 39 | 4.3 | 175287 | 70 | AP001642 |
| C | 44 | 38.8 | 4.3 | 2285 | 33 | AF170854 |
| C | 45 | 38.8 | 4.3 | 2290 | 33 | AF044000 |

ALIGNMENTS

| | | | | | | |
|------------|---------------------------------------------------------------------|----------------------------------------------------------------|-----------|----------|---------------------|-------------|
| RESULT | 1 | AB007123 | 1551 bp | DNA | BCT | 28-APR-1998 |
| LOCUS | AB007123 | Bacillus | circulans | gene for | mannanase, complete | cds. |
| DEFINITION | AB007123 | Bacillus | circulans | gene for | mannanase, complete | cds. |
| ACCESSION | AB007123.1 | GI:3090433 | | | | |
| VERSION | AB007123.1 | GI:3090433 | | | | |
| KEYWORDS | | guar gum-degrading enzyme; mannanase. | | | | |
| SOURCE | | Bacillus circulans | | | | |
| ORGANISM | | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae; | | | | |
| REFERENCE | 1 | Bacillus | | | | |
| AUTHORS | Yoshida, S. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (08-SEP-1997) to the DDBJ/EMBL/GenBank databases. Seq11 | | | | | |
| REFERENCE | Yoshida, S., Sako, Y. and Uchida, A. | | | | | |
| AUTHORS | Cloning, sequence analysis, and expression in Escherichia coli of a | | | | | |
| TITLE | gene coding for an enzyme from Bacillus circulans K-1 that degrades | | | | | |
| JOURNAL | guar gum | | | | | |
| MEDLINE | Biosci. Biotechnol. Biochem. 62 (3), 514-520 (1998) | | | | | |
| FEATURES | 88233274 | | | | | |
| source | Location/Qualifiers | | | | | |
| CDS | 1..1551 | | | | | |
| | /organism="Bacillus circulans" | | | | | |
| | /db_xref="taxon:1597" | | | | | |
| | 1..1551 | | | | | |
| | /function="guar gum-degrading enzyme" | | | | | |
| | /codon_start=1 | | | | | |
| | /transl_table=11 | | | | | |
| | /product="mannanase" | | | | | |
| | /protein_id="BAA25878.1" | | | | | |

| | | | | |
|----|-----|---------------------------------------------------------------|-----|------------------------------------------------------------|
| QY | 2 | atgcgaattccggaatttattgtaagcgggtaccactcttaacgaatgcaatgaaacccat | 61 | /db_xref="GI:3090434" |
| QY | 98 | atgcgcgaacgcgattttatgtatgacgggtaccgaattttatgacgacgaacacat | 157 | /translation="MGWFLVILRWLILAFVAFLLMFSTGQLTRKAAASGVYSGTK |
| QY | 62 | ttgtaatggaagggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 121 | LIDATGQFVNRGVNHTATWTKDLSLPIATGANTLIVLANGKRLDQVNTV |
| QY | 158 | ttgtgatggaagggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 217 | NNILILCEONKLIATLEVHDATGDSLSDILNATVNTGILASIGEDRYIYIANE |
| QY | 122 | ttgaaagggaattgcaaatccggtgtctaaatgacgttcgtttatctatgattggggagc | 181 | MYGTWDGVAMVANGKRAQAPKRLNMGVLPFTLLVDSAGMGQVDSYKNTGELNADPK |
| QY | 218 | taccagcgaattctttaaagcgtgtacgaatgcaatgcaatgcaatgcaatgcaatgcaat | 277 | NTVSIHYEYVANGKRAQAPKRLNMGVLPFTLLVDSAGMGQVDSYKNTGELNADPK |
| QY | 182 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 241 | KGVGTIAMSVMKGNSSDLATLMDTDMACNSLSTGNTVYVNSNGIKATSVLSGLFGV |
| QY | 278 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 337 | TPISSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTP |
| QY | 242 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 301 | EMKATGACTLADVLSQNSVHSYISNOMLSGKSLKATVKANNGNIGNGIYAKL |
| QY | 338 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 397 | YKIGSGGWTWYDSEENLIQSDGTILTLISGLISNLSVKEIGVEFPASSSSGSAI |
| QY | 302 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 361 | YVDSVSLQ" |
| QY | 398 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 457 | |
| QY | 362 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 421 | |
| QY | 458 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 517 | |
| QY | 422 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 481 | |
| QY | 518 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 577 | |
| QY | 482 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 541 | |
| QY | 578 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 637 | |
| QY | 542 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 601 | |
| QY | 638 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 697 | |
| QY | 602 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 661 | |
| QY | 698 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 757 | |
| QY | 662 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 721 | |
| QY | 758 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 817 | |
| QY | 722 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 781 | |
| QY | 818 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 877 | |
| QY | 782 | atggaagcaaatggaatttaacatggtgcaagcgtgtataaagacgaaggcaactatgaa | 841 | |

Query Match 46.9%; Score 422.4; DB 1; Length 1551;
 Best Local Similarity 67.0%; Pred. No. 6.2e-91;
 Matches 600; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

BASE COUNT 461 a 333 c 363 g 394 t
 ORIGIN

| DB | 878 | TGGCTATCTCGCATATGACAAATATATTGGCGGTGAACCCCTCACACCTCGTGGTAARA | 937 |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|-------------|
| OY | 842 | caatagtgaaatggtccatcatggtttaagaagaacctcgaaattaaagaccggtttt | 997 |
| DB | 938 | CCGTAGTGAATGGCAGTAACGGCATTAAAGCACTTCGTGTATTCGGCATTTTT | 993 |
| RESULT | 2 | | |
| LOCUS | D86329 | 1530 bp | DNA |
| DEFINITION | Vibrio sp. gene for beta-1,4-mannanase, complete cds. | BCI | 19-MAR-1998 |
| ACCESSION | D86329 | | |
| VERSION | D86329.1 | GI:2978250 | |
| KEYWORDS | beta-1,4-mannanase; manA. | | |
| SOURCE | Vibrio sp. (strain:MA-138) DNA. | | |
| ORGANISM | Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio. | | |
| REFERENCE | 1 (bases 1 to 1530) | | |
| AUTHORS | Tamaru,Y. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (01-JUL-1996) to the DDBJ/EMBL/Genbank databases. Yutaka Tamaru, Mie University, Faculty of Biorresource, Utilization of Marine Products, 1515 Kamihama, Tsu, Mie 514, Japan (E-mail:s9tama@maric.bio.mie-u.ac.jp, Tel:0592-31-9561, Fax:0592-31-9537) | | |
| REFERENCE | 2 (sites) | | |
| AUTHORS | Tamaru,Y., Araki,T., Morishita,T., Kimura,T., Sakka,K. and Ohniva,K. | | |
| TITLE | Cloning, DNA sequencing, and expression of the beta-1,4-mannanase gene from a marine bacterium, Vibrio sp. strain MA-138 | | |
| JOURNAL | J. Ferment. Bioeng. 83, 201-205 (1997) | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..1530 | | |
| | /organism="Vibrio sp." | | |
| | /strain="MA-138" | | |
| | /db_xref="taxon:678" | | |
| | 32..37 | | |
| | 53..58 | | |
| | 136..141 | | |
| | 149..1336 | | |
| | /gene="manA" | | |
| | 149..1336 | | |
| | /gene="manA" | | |
| | 149..1336 | | |
| | /gene="manA" | | |
| | /codon_start=1 | | |
| | /transl_table=1 | | |
| | /product="beta-1,4-mannanase" | | |
| | /protein_id="BA25188.1" | | |
| | /db_xref="GI:2978251" | | |
| | /translation="MKFTKAFSLLEPIWASCAHGEFVNSGVLYEANGSAFKINGINLAHMYTDKLSVALSGIATGANTVRVYLSNGYMTKDVSDVNIILAKANNILALAEVDTGTGEGSSASLSDSADYWEIEKNEILGEOEDVYIILNGEPGNNDAAVAMVNDHSAIORLARSAGINHTIWDADPNQGDWKEGLNNAOFVNSDPKNTLFEVHMVYVYSYNSYNDYIISSTFTNNGLVYIGERFSTRKADVDGSLMESSETLSLSTIGKSRKGNVDTSDLDIVNMNDNYSITWCGVTLINGNGIKSTSLTAVTFCGNDXNDSSGELEPCSSAVDPGDGCGWENMNSCIYVDSDTAPNGPYPCSQESSDPPDGDGWEWENNA | | |
| | MCC" | | |
| BASE COUNT | 469 a | 211 c | 333 g |
| ORIGIN | | | 517 t |
| Query Match | 26.6%; Score 239.2; DB 2; Length 1530; | | |
| Best Local Similarity | 57.1%; Pred. No. 4e-47; | | |
| Matches | 521; Conservative 0; Mismatches 373; Indels 18; Gaps 4; | | |
| OY | 3 | tgcaaatccggaatttatgttaagcgggtacacactctatcgatgccaatggaaccatt | 62 |
| DB | 202 | TGCTATATCGGGGTTTATATGATCAAAATGCTCTTATATGAAGGAAAGTAGTCCCT | 261 |
| OY | 63 | tgtaatgaaggaattacaatcgtggcagcgcgtgataaagcagcagcaactcgaat | 122 |
| DB | 262 | TAAATTTAGAGGAGTAATATCATGCACATACCTTGTTATACAGCAAGTTTAGCGTGGCTTT | 321 |
| OY | 123 | tgaaggagattgccaataaccggtgccaataacggtccggaattggttatctgatactgaatgggagaca | 182 |

Db 322 AAGTGTATGTGTCGCCACAGAGCCAAATACGTACGGGTTTACTAACTAAATGGTTATCG 381
Oy 183 atgacaaagatgatgacatccacacagtaagaacacttaccctttagcggagaataatca 242
Db 382 GTGGACAAAAAATAGTACGTTAGTGAATGTACATCAACTTAAATTTAGCAAAAGCAACAA 441
Oy 243 ttgggttcgtctcttgaagtcataatgatctaccggtaatg-----atccaatgc 293
Db 442 CCTTATACCTATATATAGAAATTCATGACACACTACCTGGGATATGAGNAGAAATAGTGTGC 501
Oy 294 ttgcgtcaatcgttgtgtgtattatgtgattgaatggaatggaagtgcttaatttgaaagga 353
Db 502 TAGTTTAAGATTCCTGGCGGTGATTAAGTGAATGAGTTGAGATGACCTAAATCGGCAACGA 561
Oy 354 agataccgtcatatataattgacgaagaatggtgtg----ttcgtggaaggagatgc 410
Db 562 AGATTATGTCACTCAATTAACCTAGGCAATGAACCTTTTGATATATATATGATGGCGGTAC 621
Oy 411 ttgggttcacgggtataaacaagaacatcccgagatctgcgttaacgcccgttcaaacatcac 470
Db 622 TTGGTATATGACCATGTCAGTGCATTAACAAGATTAAGAAGTGAAGGTATCAATACAC 681
Oy 471 ctgtgtgtatagatcgtcgtgggtgtggggacaa---ttccacaatcgatcatgattagg 527
Db 682 GATTATGTGTTGATGCTCCAAACTGGGGCCAAAGACTGGAAGGCTTTATGCTGAATTAATGC 741
Oy 528 aagagaagtttttaattgctgacccccaacgaatacaatgttttcgcatcatatgata 587
Db 742 TCATATTCGTTTAAATTTCTGACCCCTAAATTTGAATTCGATTTTATGCTCATATATATGA 801
Oy 588 atatgcagggtgtaatgatcgcgaagttcgtactaataatgaccgagttccttaataaga 647
Db 802 GGTATATAGTATGATCAATTC---AGTAAAGACTACATCTCATCATCATTTACAAACAAATGG 858
Oy 648 cctgcgcatatgcatatggttgaatttgacacccgtcatatacaaatgttgcgcgatgaagc 707
Db 859 CTATAGTGTGTTACGGGTGAGTTCGCTTCACTCATTAAGAGGGCCGATGTGATGAAGG 918
Oy 708 aacggtttgagctatcttgacaaaggagggttggtgtgtgtgctgtgcatatgaaagc 767
Db 919 TTCAATTAATNGACAGCTTAGAAACCTTAAGTTTAAAGTTATATTCGGTTGTGCTGTACGG 978
Oy 768 gaacggcccaaatgtagtgagatttagaaccttcgcaatgatgtaggcctggaataaactctac 827
Db 979 AAATGATATCTCAACATCTCGAATTTGGAATATGTATATTAATTTGGGACATATTTGCTATAG 1038
Oy 828 agcttggggaaaatacaataatgtaatgtagtgcataatgatttgaagagaacttcagagattaag 887
Db 1039 CACTTGGGGGTATGTTTAAATTAATGATGCTCAAAATGGTATTAATCCATCATGACAGTTTACG 1098
Oy 888 caccgtttttac 899
Db 1099 TACGGTGTTTAC 1110

RESULT 3
AF163837
LOCUS AF163837 4567 bp DNA BCT 08-FEB-2000
DEFINITION Caldicellulins cellulovorans multidomain beta-1,4-mannanase
precursor (manA) gene, complete cds, and unknown genes.
ACCESSION AF163837
VERSION AF163837.1 GI:6651325
KEYWORDS
SOURCE Caldicellulins cellulovorans.
ORGANISM Caldicellulins cellulovorans
Bacteria; Firmicutes; Bacilli; Clostridium group;
Bacillus/staphylococcus group; Caldicellulins.
1 (bases 1 to 4567)
Sunna, A., Glibbs, M.D., Chiu, C.W.J., Nelson, P.J. and Bergquist, P.L.
A gene encoding a novel multidomain beta-1,4-mannanase from
Caldicellulins cellulovorans and action of the recombinant enzyme on
kraft pulp

[illegible]

| Query Match | Best Local Similarity | 14.9% | Score 134; | DB 1; | Length 4567; |
|-------------|-----------------------|--------------------------------------------------------------------|------------|-----------------|--------------------|
| Matches | 408; | Conservative | 0; | Mismatches 360; | Indels 18; Gaps 4; |
| QY | 68 | tgaagaggaataacacatgycacgcacatggtataaagacacgaactactgcaattgaag | 127 | | |
| Db | 2070 | TGTTGGTATTAAACACCCCGCATCGTGTATCCGACGGCGCTCAGTTCGTCGTTGCAAG | 2129 | | |
| QY | 128 | ggattgcacaataacacggytgtaataacggtccggaattggtatctcgtatggyggaacaatgga | 187 | | |
| Db | 2130 | GCATTCGTTCTCTGGGGGCCACACGGCGGCGGATCGCTGAGACACGGTTGGCCGGTGA | 2189 | | |
| QY | 188 | canaaagatgacatcatcatcagtaagaaacactatctctttagcggaaagataa---tcatt | 244 | | |
| Db | 2190 | CGAAGATCCGGCCACGCGCAAGTAGCGGACATATTTCACAGCGCGGACGCTCGGTATC | 2249 | | |
| QY | 245 | tggttgctgctcttgaagatcattgactacggtttatgattccattgc-----ct | 295 | | |
| Db | 2250 | GGGGGGTCTCTTGAAGTTCCACACACGACCGGCTACGGGGAAGACGGCGCGCTGTT | 2309 | | |
| QY | 296 | cgctcaatcgcgcgttgattatgattgattgaattgaagagaagggcttaattggaaggaag | 355 | | |
| Db | 2310 | CGATTGACGACGGCGGTCAATTATTGGATTGCACTGAAGAACGTCTGCGGGTTAGAGGA | 2369 | | |
| QY | 356 | ataccgtcatattaatattatattgycgaattgaattggttctgtyggaagggatgcttggg | 415 | | |
| Db | 2370 | ATTTCGTCATCGTCAAATATCGGCATATGACCGGTACGGGGAACAATATTATCAGAACTGGG | 2429 | | |
| QY | 416 | ctgacgggtataaacaagaagaatcccggaattggttaacgcgcgttctaacaactacttga | 475 | | |
| Db | 2430 | TGACGCGACACCGCGGAACGGGTGACGCGCTGCGGAAACGGGGGATTAACAATATCGATCA | 2489 | | |
| QY | 476 | tggtatagatgctgcggygttggggac---aatccacaatcggatcatgactatgaaagag | 532 | | |
| Db | 2490 | TGTTGGACGCTCCGACACTGGGGGTGAGGACTGTGCTTCACAGATGCGGGACACGGCGCGA | 2549 | | |
| QY | 533 | aagttttaatgctcgaccctccaaagaataacaatgcttttcgatctcataatgataatg | 592 | | |
| Db | 2550 | CGATTTTCACAGCGCGACCCGACAGAGAACTGGCTTTTAAATCCACATATTACGGG---G | 2606 | | |
| QY | 593 | caagttgtaagcaatcgcaagttcgtctaatttgaacggaattcttaacaaagacctcg | 652 | | |
| Db | 2607 | TGTACGACAGGGCGCGGAGTGAAGATTATATGTAGTGCTTCGTATAACCGCGGTTGC | 2666 | | |
| QY | 653 | cattatgcatatgtaatttggaacacgcgtcatacaaatggtgacgtcatgataagcaaga | 712 | | |
| Db | 2667 | CGCGTGAATCGGGGATTTTGTATATGACACGCGACGGGATTCGACAGCAGCGCA | 2726 | | |

| FEATURES | source |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Db 2727 | TCGTACATATATCGAAGCAATCAACATTGGCTTGTTCGGCTGGTCGTGGACGGTAACG 2786 |
| Db 2787 | GAGCGGGCGGTGAATATTGTGACATGTCGACAACTTCAATCGAAGACCCGAGCGGT 2846 |
| Db 2847 | GGGGGA 2852 |
| RESULT 4 | |
| LOCUS | CDCCMANABD 2155 bp DNA BCT 26-APR-1993 |
| DEFINITION | C.saccharolyticum beta-D-mannanase (mana) gene, complete cds. |
| VERSION | M36063.1 M37147 |
| KEYWORDS | M36063.1 GI:144292 |
| SOURCE | 1,4-beta-D-mannan mannanohydrolase; beta-D-mannanase. |
| ORGANISM | C.saccharolyticum DNA, clone lambda-WP2. |
| REFERENCE | Caldicellulosiruptor saccharolyticus |
| AUTHORS | Bacteria; Firmicutes; Bacillus/Clostridium group; Thermocaeobacter group; Caldicellulosiruptor. |
| TITLE | 1 (bases 1 to 2155) |
| JOURNAL | Luehti, E., Jasmut, N.B., Grayling, R.A., Love, D.R. and Bergquist, P.L. |
| MEDLINE | Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium Caldocellum saccharolyticum |
| COMMENT | Appl. Environ. Microbiol. 57, 694-700 (1991) |
| | 91247819 |
| | Draft entry and computer-readable sequence kindly submitted by E. Luehti, 03-JUN-1990. |
| | Author address: E.Luehti |
| | University of Auckland |
| | Department of Cellular and Molecular Biology |
| | Private Bag |
| | Auckland |
| | NEW ZEALAND |
| | email:SYSTEM@CELVAX.AUKUNI.AC.NZ. |
| | Location/Qualifiers |
| | 1. 2155 |
| | /organism="Caldicellulosiruptor saccharolyticus" |
| | /db_xref="taxon:44001" |
| | <1. 685 |
| | /codon_start=2 |
| | /transl_table=11 |
| | /label=ORF1 |
| | /protein_id="AA072860.1" |
| | /db_xref="GI:144293" |
| | /translation="GSNWTFGFQAMSHQVVAEYYVTGDKDAGALLKRWVMIKSYVW |
| | LNSGTFAPISPTLDMSGQDPDNGETTYGPNLHYKVYDYGTDGITSALVALYYSAS |
| | GTRKYGVDEPAKRLAKELIDRMKIRVDEGLSAPERADARYKFFDEYVITPGMTG |
| | KMPGVDYKSGVKTIDIRSKTKQDPMPKLEAAYKSGVPEFRHREMAQCDIAIVNA |
| | TYELFSGNQ" |
| | 818. 830 |
| | /note="mana ribosome binding site" |
| | 841. 963 |
| | /product="beta-D-mannanase" |
| | 841. 1881 |
| | /EC_number="3.2.1.78" |
| | /codon_start=1 |
| | /transl_table=11 |
| | /product="beta-D-mannanase" |
| | /protein_id="AA072861.1" |
| | /db_xref="GI:144294" |
| | /translation="WRLEKTKIRKRWLSYLCTVVFLLNTLFLTANTYILPKGCAATSDNG |
| | VVKIDTSLIGTINAHGWCYRDRLDTALGISWGNMSVRYVLSNGYRWTKIPASEVNA |
| | IISISRSFGFAIILLEVHDTGYGEDGNAAGSLAAVEWKEIKSVLDGNEDFVILNIG |
| | NEPQNNNYQMWVNDTKNAIKALRDAGCKRIITMDAPWQGDMSMTMDNDSITMELAD |
| | PLRNLVYSIHMYGVNTSKVEEYIKSPVVDGDLVYGEFGHHTDGDQDEDAIVRYR |
| | KQYKIGLFMSWMCNNSVYGLDMAVNNMDPNPPIPMGWMTYITNALIGTSITPTPISTYV |

| Query Match | Best Local Similarity | Score | DB 2: | Length | 2155: |
|-------------|---------------------------------------------------------------|-------|------------|--------|--------|
| Matches | 407; Conservative | 0; | Mismatches | 362; | Indels |
| | | | | | Gaps |
| 68 | tgagagagattaacctgagcagcagctatgaagaccagcgaactatgcaattgaag | 127 | | | |
| 998 | TAATAGAACCAATCACGCACATGCTGTTGACAGAGATGACTTGTACGGCATTCGGC | 1057 | | | |
| 128 | ggattgcaaatccggtgctaatacagtcggaattggtatctatctgagggagcaatga | 187 | | | |
| 1058 | GAATTAGTCATGGGGGTATGAACTGCTGTGAGGGTACTGCTGATGATGTTACCGATGA | 1117 | | | |
| 188 | caaaagatgacatccataaagaagaacctatctcctttaacggaagataatcattgg | 247 | | | |
| 1118 | CGAAGATACCACCAAGTGAAGTGGCCAAATTTATATCATTTGTCAGAACCTCTGGTTTA | 1177 | | | |
| 248 | tcgcgctt---cttgaagtcatagtatgcaccggtatgattccattg-----cct | 295 | | | |
| 1178 | AAGCTATTATTATTAAGTGCACGACACAAAGATATGAGAAAGATGGGCGACATGTT | 1237 | | | |
| 296 | cgctcaatcgtcgttgaattatgattgaaatgagaagtccttaatttggaaaggaag | 355 | | | |
| 1238 | CATTGGCACAACAGATGGAAATTTGGAAAGAGATTAAGAAGCGTATTAGACGGTAACGAAG | 1297 | | | |
| 356 | ataccgcatcttaataatctcgaatgaatggttggctcgtgggaaggagatgctggg | 415 | | | |
| 1298 | ATTTTGTAATTTTAAACATTTGTAATGAGCCGTATGAGAACATATCAATCAAACTGGG | 1357 | | | |
| 416 | ctgaagcggtataaagaagatcccgagctgctgaagcgggtctaaaccatacttga | 475 | | | |
| 1358 | TTAAATGACACGAAAGACGCTTATTAAGCCTTAGAGATCAGAGATTCAGCACACATTA | 1417 | | | |
| 476 | tggtagatgctcggggtgggagc---aatttcacaatcgaattcagaatttgaagag | 532 | | | |
| 1418 | TGCTGGATGCGCGGAACGTGGGCTCAGAGATTTGCTTAATCTATGAGAGATATGCCACA | 1477 | | | |
| 533 | aagctttaaagctgaacctcaagaagaatacaatgcttcgaattcaatgataatg | 592 | | | |
| 1478 | GCATATGGAACACAGATCCGCTGCCCATTTGTTTGGATTCATATATGATG---GCG | 1534 | | | |
| 593 | caggtgtaatgcatcgcaagctcgtaacatataatggaacgaggtcttaacgaacctg | 652 | | | |
| 1535 | TATTAATATACCAACAGCAAGTCCAGAGATACATCAATCATTTGTTGAAGGGTTAC | 1594 | | | |
| 653 | catgaatcatgtaatttgacacccgcatcacaaatggtgaagctgcataagcaaga | 712 | | | |
| 1595 | CATTGCTTATTTGGAGATTTGGACATCACACACAGATGTACCTGATGAAGAAGCTTA | 1654 | | | |
| 713 | ttaatgactatctcaacaagaagagttggttggctcgtgctcatgagaaaggaagc | 772 | | | |
| 1655 | TTGTCAAGTATCAAAACAGTACAAAGATATGAGATTTATTTGTTGGTCTGTGGAAT | 1714 | | | |
| 773 | gcccaagaatggaagatttagacccttcgaatgattggctcgtggaataacccttaagct | 832 | | | |
| 1715 | CGAGCTATGTTGGTATTTGGACATGATGTAACAACTGGGACCCCAATATATCCAACTCAT | 1774 | | | |
| 833 | ggggaag 839 | | | | |
| 1775 | GGGGACA 1781 | | | | |

GRVVEGNSAFVMEVNHAAVWYDPDRGTSTADIANKANTRVYLSSGGKWTSE
 VSALIGCKRANKVICYLEVHTTTCYGDGATSLQAGIYWGVSAMRAREDVYVN
 IGFEPGNTAANTDARKSAIGLRGGLGHALMVDPMNGQMSGMSNLSVTA
 SDPDRNTVFSIHMTGYDITAEVBDYPLNAFNGNLPIVGEFGQSDGNDDEDAIVA
 TAOISIGVYGLMWSNGNGVYEDLDNMGSDPNSILTSNGNILLGSGNATSTATVY
 YGGGGSTGGTAPNGYPCVYNGASDPDGDGMWENSRVCSVRSADH"

sig_peptide 226..333
 /gene="mana"
 mat_peptide 334..1374
 /EC_number="3.2.1.78"
 /function="hydrolase"
 /product="mannanase"

BASE COUNT 309 a 648 c 647 g 277 t

Query Match 9.5%; Score 85.8; DB 74; Length 1881;
 Best Local Similarity 47.5%; Pred. No. 1.9e-10;
 Matches 425; Conservative 0; Mismatches 452; Indels 18; Gaps 5;

13 ggattttatgaagcggtaaccctctatcgatgccatgaaaccatttgtaagaga 72
 340 GGCATCCAGCTGACAGAGGAGCTGCTCGAGGCAAGCGGCGCTTCATATGGC 399
 73 ggattaccatggcagcagcatgataaagcagcactactgcaattgaaaggatt 132
 400 GGGGTCAACACGCTACCTGATATCCCGAC---GCACGGGCTCCATGCCACATC 456
 133 gcaaatccggtgtcaatacggctcggatgtgtatctgattgaggagacaanaa 192
 457 GCGGCAAGGAGGACCAACCGTCCGCGCTGCTACAGCAGGCGGCGCTGAGCAAG 516
 193 gatgacatccatacagtaagaaccttctcttaagcgaagaataatattgttgt 252
 517 ACGAGCGGCTCCGAGGCTCTCGGCTCATCGGCCAGTGCACAGGCCACAGGTCTGT 576
 253 gtcttgaagttcatgattgtacgggttaattccattgcttcgtca-----atcgt 306
 577 GTCTTGAAGTGTACAGACACCGGCTACGGCAAGAGCGGCGCAACCGCTGAGACG 636
 307 gctgttatttattgattgaatgaagaagtgtct--ttaattgaaaggagaataccgtc 363
 637 GCGGCGGACTACTGCGGTGCTGAAGACCGCGCTGAGGCGCCAGAGGAGACTACGTC 696
 364 attataatattgcaatgaagaatgtgtgtcgtggaaggagagatggtggtcagg 423
 697 GTCTTCAACATCGGACAGCAACCTTCGGCAACCACTACGCGCGTGGACGAGCGG 756
 424 tataaacaagaacatcccgatgtgttaagcgggtctaaaccataccttgaagtag 483
 757 ACCAAGTGGGCAATCGGCAAGCTCCGCGGCGCGCTGCGCCAGCCTGATGCTGAC 816
 484 gctcgggggtgggagc--aatcccaaatgattcatgattatgaaagagaagtttt 540
 817 GCGGCCAATCGGCGGAGAGACTGTCGCGGCAAGTCCGCGTCCAAAGCGCGCTCGGTCCTC 876
 541 aatgtgcccccaagaaataatgatttgaattcatatgataatcaagtgtg 600
 877 GCCTCGAGACCCGACGCAACCGCTCTTCTCATCTCACTATGAC--GGGTCTACGAC 933
 601 aatgatacgaagttcgtactaataattgacgagttcttaacaaagacctcgaattg 660
 934 ACCGCGCGCGAGGTGGGGACTACCTGAACGGCTTCGTGGGCAAGGAGATGCCATCGTC 993
 661 attgtgaattggaacacgctcatcaaatgtgaagtcgtaagaaagcaagattatgagc 720
 994 GTCCGCGAGTTGGGAGACAGCACAGTACGCGCAACCCGAGAGAGACCCCATCATGGGCG 1053
 721 tattctgaacaagaagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
 1054 ACGGCCAGTCCCTCGGCGCTGCTACCTCGGCTGCTCTGAGGCGGCAACGGCGGCGGC 1113
 781 tggaggtatttagaccttcgaatgattggtgtgaataaccttcaagcttgaggaaat 840

1114 GTCGAGTCCCTCGACATGCTGACAGGCTTCGACCCGAACTCGCTGAGAGAGCTGGGGCAAC 1173
 841 acaatagtaatgtgtccatgattgatttaagagaaccttcgagattaaagccgttt 895
 1174 GCGATCCTCTACGCGAGCAACGCGATCGCCGAGCTCCAGGAGCGGCGCACCGCTGT 1228

RESULT 9
 TFU6227 837 bp DNA BCT 03-DEC-1998
 LOCUS
 DEFINITION Thermomonospora fusca man gene, partial.
 ACCESSION AJ006227
 VERSION AJ006227.1 GI:3970819
 KEYWORDS beta-mannanase; man gene.
 SOURCE Thermobifida fusca.
 ORGANISM Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae; Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.

REFERENCE 1 (bases 1 to 837)
 Hllge,M., Gloor,S., Winterhalter,K. and Piontek,K.
 Direct Submission
 Submitted (18-MAY-1998) Hllge M., Biochemistry, Federal Institute of Technology Zuerich, Universitaetsstr. 16, CH-8092 Zuerich, SWITZERLAND

REFERENCE 2 (bases 1 to 837)
 Hllge,M., Gloor,S., Winterhalter,K. and Piontek,K.
 Crystalization and preliminary crystallographic analysis of two beta-mannanase isoforms from Thermomonospora fusca RM3 Acta Crystallogr. 52, 1224-1225 (1996)

REFERENCE 3 (bases 1 to 837)
 Hllge,M., Gloor,S.M., Rypniewski,W., Sauer,O., Helgman,T.D., Zimmermann,W., Winterhalter,K. and Piontek,K.
 High-resolution native and complex structures of thermostable beta-mannanase from Thermomonospora fusca - substrate specificity in glycosyl hydrolase family 5 Structure 6 (11), 1433-1444 (1998) 99036862.

TITLE JOURNAL MEDLINE
 FEATURES
 SOURCE Location/Qualifiers
 1..837
 /organism="Thermobifida fusca"
 /strain="RM3"
 /db_xref="taxon:2021"
 1..837
 /gene="man"
 <1..>837
 /EC_number="3.2.1.78"
 /function="hydrolase of mannan"
 /codon_start=1
 /trans_table=1
 /evidence=experimental
 /product="beta-mannanase"
 /protein_id="CAA06924.1"
 /db_xref="GI:3970820"
 /translation="GLHVKNGRIYEANGOEFTIRGVSHPHNWYPOHTOAFADIKSHCA
 NTVRYVSNQVRSKNGSPDVAIVISICRKNRLICMEVDHTGSGSGASTLDQV
 DYRIELKSVYQGEEDYLYLINIGNEPYNDSATYAAAGWDISAQIRARAGFEHTIV
 DAPRWGDWNTMRNNADQVYASDPSTNTVFSIHMTGYDITAEVBDYPLNAFNGNLPIV
 LIIGEPHDSHDGNDPDDITMAERLKLIGYIGWSNGNGGVYEDLDNMGSDPNSILTSNGNILLGSGNATSTATVY
 PWG"

BASE COUNT 169 a 292 c 255 g 121 t

Query Match 9.1%; Score 81.6; DB 74; Length 837;
 Best Local Similarity 48.0%; Pred. No. 2e-09;
 Matches 404; Conservative 0; Mismatches 414; Indels 24; Gaps 5;

13 ggattttatgaagcggtaaccctctatcgatgccatgaaaccatttgtaagaga 72
 1 GGGGTCAATGTAAGAACGCGCGCTGTATGAGGCCACGCGGAGACTCATCATCGCT 60

```

Oy 73 gggattaaacatgagcagcatggtataaagaccaggaacactactcgaattgaaggatt 132
Db 61 GCGGTCAGCCACCCCAACTGT---ACCCCAACACCCAGCCGTCGCCGACATC 117
Oy 133 gaaataaccggtcgaataacggtccggtatgtgtatctcgtatgggggagaatggaaana 192
Db 118 AAGTGGCAGCGGGCCCAACCCGTCGGGTGTGCTGAGCAAGGTTCTCCGGTGAACAAAG 177
Oy 193 gatgacatccatacagtaagaacactatctctttagcggaagataatcatttggttgt 252
Db 178 AAGGCGCTTGTGAGAGTGGCCCAACGTCATCTCCCTGTGCAAGAGAACGCCCTATCTGC 237
Oy 253 gttcttgaaatcagatgagtcacgggtatgattccatt-----gttcgtcgaat 303
Db 238 ATGCTGGAGGTGACGACACCCGCTACGGTGAAGCAGAGCGGGCCCTCCAGCGTCGAC 297
Oy 304 cgtgctgttattgtattgattgaatgagaagtgcttattattggaaggaataaccgfc 363
Db 298 CAGGCGGTGACTACTGATGATCGAGCTGAAGAGCGTCTCCAGGGCGAGAGGCTATGTGC 357
Oy 364 attataatattcgaatgaatgattgtgtt-----tcgtgggaagggtatgtgtgct 417
Db 358 CTATCAACATCGGAGAGAGAGCCCTACGCGCAAGACTCCGCGACGCTGCCGCTGGGGG 417
Oy 418 gaagggtataaagaacgaatcccgagattgctgaacgctgttaaacataccttgatg 477
Db 418 TGGGACACCTCCGCGCCATCCAGCGCGCTGCGCGCGGATTCGAGCACACCTCTGTCG 477
Oy 478 gtagatgtcgggggtggggagc---aattcacaatgattcatgattatggaagana 534
Db 478 GTGGAGCCCCCACTGGGGGCAAGACTGAGAGAACCATATGGGGAACAAAGCCGACCAAG 537
Oy 535 gtttaaatgctgaccccaacgaataacaaatgatttgcattcatatgatatgata 594
Db 538 GTGTACGCCAGGAGACCCCAACCGGCAACACCGCTTCTGATCACAATGTACG---GCGTC 594
Oy 595 ggtgtaatgatacgcgaagtcgtactaaatattgaccggaattcttaataagaacctgca 654
Db 595 TACTCCCAAGCGCTCCACAGATCACAGCTACCTGAGACCACTTGTCMAACCGCGGCTCGCG 654
Oy 655 ttagcatattggtgaatttggacaacgctcacaataagtgatgataagaagatt 714
Db 655 CTATCATGTGGGAGTTCGGCCACGACCACTCCGAGCGCAACCCGACAGGACAGATGATC 714
Oy 715 atgagcatattcgaacaagaagagtggtggtggtgctgctgcatggaagaagcgcc 774
Db 715 ATGGCCGAGAGCCGAGCGGCTCAAGCTGGGCTCATGCGCTGCTGAGAGTGGCAAGCGC 774
Oy 775 ccagaatggagattattagaccttgcgaatgattgggctggaataaccttcaagcttg 834
Db 775 GCGGCGGTGACTACCTCGACATGTGTACAACTTGACGCGGACGACAACTGAGCCGCTGG 834
Oy 835 gg 836
Db 835 GG 836

```

```

RESULT 10
SCF73/c 14922 bp DNA BCT 04-OCT-1999
LOCUS Streptomyces coelicolor cosmid F73.
DEFINITION AL121746.1 GI:6013076
VERSION AL121746.1 GI:6013076
KEYWORDS catalase/peroxidase; cpeB; Fe regulatory protein; furs; manA;
membrane-bound oxidoreductase; MerR-family transcriptional
regulator; polyphenyl synthetase; secreted beta-mannosidase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 14922)
AUTHORS Redenbach, M., Kleser, H.M., Denapalte, D., Eichner, A., Cullum, J.,

```

TITLE Kinashi, H. and Hopwood, D.A.
 JOURNAL A set of ordered cosmids and a detailed genetic and physical map
 MEDLINE for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 97000351
 REFERENCE 2 (bases 1 to 14922)
 AUTHORS Seeger, K.J. and Harris, D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 14922)
 AUTHORS Parkhill, J., Barrell, B.G. and Randles, M.A.
 JOURNAL Direct Submission
 TITLE Submitted (04-OCT-1999) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Notes:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh1.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
 Cosmid F73 overlaps with cosmid F11 on the AseI-F genomic restriction fragment.
 Location/Qualifiers
 1..14922
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid F73"
 1..122
 /note="nominal overlap with Str11"
 complement(1..983)
 /gene="manA"
 complement(1..983)
 /gene="manA"
 /note="SCF73.01c, manA, probable secreted beta-mannosidase, len: 327 aa; similar to many e.g. MANA, STR11 P51529 manan endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189 2-score: 1332.7 E(-): 0, 61.1% identity in 288 aa overlap. Contains N-terminal signal sequence, and Pfam match to entry PF00150 cellulase, Cellulase (glycosyl hydrolase family 5). Also similar to SC2H4.16 (62.2% identity in 288 aa overlap)"
 /codon_start=1
 /transl_table=11
 /product="putative secreted beta-mannosidase"


```

/protein_id="CAB57406.1"
/db_xref="GI:6013077"
/translation="MRPARDIPIRTIPARLLIGLALIGLVVVGALCPGALAQSPAA
RSAGVSAQAAGLHIGGRLLGNGNFMVGMVHARTWPGETQSADYALGAN
SVYVILSDIRMSNGPADAAYAEQCKANRLICVLEVDITGVAADAAQTLIDHAD
YVIGLVDVAGDEYIVINIGNEPMGNTDPAWTEPTVAVKRLAAGAOHIMVADP
MNGDQGVARARASVYDADPTGNLIFSIHMSVFDTQETIDYANAVDELPIILI
GERGPDADQGDDEDDMMATLQGLRLGLIAMSWSGNTDPIVDLALDDPSS"
RBS
gene
complement(989..992)
complement(1129..2724)
/ gene="SCF73.02c"
/ complement(1129..2724)
/ note="SCF73.02c"
/ protein_id="CAB57407.1"
/ protein_id="CAB57407.1"
/ db_xref="GI:6013078"
/ translation="MSDHDKTPNPSARFRTWNLPGALSGLLAGALAAVLAIAVR
PRSGVAVVAGGAIDRTPTAVKDMALRTGIDKVLQIGIILVLELALGAAGR
HRRAGAVLVGIVGAALAGIPDSTGYTDAFSPVLAGAVILLYLLGLLPDPR
EAPGERGEMDRRRFVIATAAASAGTGVGRALSGAGREAVASRENTLPPDS
RAPVPROARAVEGVSFTPTSDPYRTDLYPKVDTATMLRIRIBEGVARRTL
TDDILRRKLIERDITLTCVSNVEGSPYGNARWIGVRLADLACGRAPSGGPAD
QIVASVDTMTIGSPVEDWDSRDALVANGSEPLPEDHGFVARRVVGILGYSAQ
KNIETDELTFDSYDAVYVKRWARAPVKTSLDTPFPFAPRAGVYVAGVMAQ
HRGIDKVEVRVDDGEVAVLAEDSRDTQWMSYAMRATKGHITLVATDRTGEVO
TDRKRTVVDGASGRHSVYVTD"
/ gene="SCF73.03c"
/ complement(2777..4300)
/ protein_id="CAB57408.1"
/ protein_id="CAB57408.1"
/ db_xref="GI:6013079"
/ translation="MTAFILVSGMFTGHIWODTARLARGAEVHTVATLGLDAPRA
AAAPVDLETHADYLAVIDSAGADRIVYVGDYGHIPAVGAADRRAERVDIVH
LDSGLPBDVPALAAVPOSLERLAGAAGADVLPRAHEMPMGSTAGVPODALD
RLTALAAPOLGTLIOPRLTGAADVPTPTGVLGNGNSTEILYOMVRLDPRALRL
TDPVRSFEIIPRGHWMPSTIPELIVDYLIRAAAGGRKLEPDDTAGHGLHPLMDY
PDVPRHRGNLDITPDAEPRPAVVLVHGSPVADPARDTPMDWGLGYACAVAGDG
AVGALLDRLHDUSFERAADVAVAEVRADPDRDALVDFSSGGGLIADWLD
AEPAMRLCAATPYLALPMMGLSESEFRPRAVANGALVLTVRGRENPELTAAT
VEEFLAAKDCADVEVDVPRGHGFEITDLDTSRAVRAALTVLDHAFGAR"
/ gene="SCF73.04"
/ complement(4311..4315)
/ protein_id="CAB57409.1"
/ protein_id="CAB57409.1"
/ db_xref="GI:6013080"
/ translation="MFTSNDIGCGIGELAEAGVYVKTVPREYSDRGLLEPSSRSAGH
RNYGEALDRRLRLRLIGLIGVEVRRITDERNAHSTIAGEBGALEAAVAGRLREV
GSEALALMRREALRLVDEAPPGEDADRLRLIGAVSPSTAPLVKRWGLPAPMPA
RSAGFLVAVPQPEDEPAPQAVLAFAALNMLTAPCSPGTVQDQPEHRAAGAA
LLYGLAEVLAELAGVHMRGRPEHGEALDAFVASYAVGRTIPGRRLAAGLAA
DPRLDGWELVAEVTTAPGARPEPTPGSAHMLLAALTAGAVDDQTSPOSKRRV
SPGLGV"
/ gene="SCF73.05c"
/ complement(5357..5623)
/ protein_id="CAB57410.1"
/ protein_id="CAB57410.1"
/ db_xref="GI:6013081"
/ translation="MDEVEMRGVYVYADHNDPGRPRAGRVYVHLVGGPLDGLLVNTE
LVEQERAGVTLATEIGRYGPGHSAVTPRPDTRRDMRDVP"
/ gene="SCF73.06c"
/ complement(5697..6697)
/ protein_id="CAB57411.1"
/ protein_id="CAB57411.1"
/ db_xref="GI:6013082"
/ translation="MLNDTPVTVDGRGRTDVTWMEPAMRTEAIGIEARLAAGLR
PTGRSVRLRPMVSLRLCVAGDDVAWFKSPPGAEALATASIAWVPHVLEP.A
VDAARGLSLDEGDLPLFDVLAAREVPGDEEILROYAAMHGTSRASIEGLVY
AVPAGAOGLVDRLDTALSRADRALROLRPLLDIMCAELTALGVDDSLDHADLHG
OLFPARGREFEFMGDAVYVQPSLSAVPARRAAREVGYOVDRLDAYLEPTGGS
RTAOPRLRAVSLARLRLNRAAYIGRLPFAASGASAMATRAARCLLELDEFTI"
/ gene="cpeb"
/ complement(6778..9000)

```

```

/transl_table=11
/product="putative MERR-family transcriptional regulator"
/protein_id="CAB57409.1"
/db_xref="GI:6013080"
/translation="MFTSNDIGCGIGELAEAGVYVKTVPREYSDRGLLEPSSRSAGH
RNYGEALDRRLRLRLIGLIGVEVRRITDERNAHSTIAGEBGALEAAVAGRLREV
GSEALALMRREALRLVDEAPPGEDADRLRLIGAVSPSTAPLVKRWGLPAPMPA
RSAGFLVAVPQPEDEPAPQAVLAFAALNMLTAPCSPGTVQDQPEHRAAGAA
LLYGLAEVLAELAGVHMRGRPEHGEALDAFVASYAVGRTIPGRRLAAGLAA
DPRLDGWELVAEVTTAPGARPEPTPGSAHMLLAALTAGAVDDQTSPOSKRRV
SPGLGV"
/ gene="SCF73.05c"
/ complement(5357..5623)
/ protein_id="CAB57410.1"
/ protein_id="CAB57410.1"
/ db_xref="GI:6013081"
/ translation="MDEVEMRGVYVYADHNDPGRPRAGRVYVHLVGGPLDGLLVNTE
LVEQERAGVTLATEIGRYGPGHSAVTPRPDTRRDMRDVP"
/ gene="SCF73.06c"
/ complement(5697..6697)
/ protein_id="CAB57411.1"
/ protein_id="CAB57411.1"
/ db_xref="GI:6013082"
/ translation="MLNDTPVTVDGRGRTDVTWMEPAMRTEAIGIEARLAAGLR
PTGRSVRLRPMVSLRLCVAGDDVAWFKSPPGAEALATASIAWVPHVLEP.A
VDAARGLSLDEGDLPLFDVLAAREVPGDEEILROYAAMHGTSRASIEGLVY
AVPAGAOGLVDRLDTALSRADRALROLRPLLDIMCAELTALGVDDSLDHADLHG
OLFPARGREFEFMGDAVYVQPSLSAVPARRAAREVGYOVDRLDAYLEPTGGS
RTAOPRLRAVSLARLRLNRAAYIGRLPFAASGASAMATRAARCLLELDEFTI"
/ gene="cpeb"
/ complement(6778..9000)

Query Match      5.8%; Score 52.4; DB 2; Length 14922;
Best Local Similarity 46.9%; Pred. No. 0.017;
Matches 275; Conservative 0; Mismatches 266; Indels 15; Gaps 2;

QY 11 ccgagattatgtaagcggtacacactctatagatgcaatggaaccattgtaatga 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 814 CCGGCCCTGCACATCGGTACGCGCGGCTGTGAGGCAAGCGACATCGTCATGC 755

QY 71 gaggattacacatggtgacgcattgataagaccaggaactactgcaattgaagga 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 754 GCGGGGTCAACACCCACACCTGTGTATCCCGGAGACGCACTGCTGGCGGACGTCA 695

QY 131 ttgcaataccggtctataacggtccgattgtgtatcctgtaggggagacatgacaa 190
    || || || || || || || || || || || || || || || || || || || ||
DB 694 AGCG---GCTGGGGCGGAACAGCGTGGTCTCTCCGAGGGGACCGCTGGAGCG 638

QY 131 aagatgacatccatcacgtgaagaacctatctcttagcggagaataatcattggtg 250
    || || || || || || || || || || || || || || || || || || || ||
DB 637 AGAAGGCGCCGCGGACGCTCCCGCGCTCATCGACATGCAAGGCGACCGGCTCATCT 578

QY 251 cgttcttggaagtcatactgctacgggtttgattccattgc-----tgcgtca 301
    || || || || || || || || || || || || || || || || || || || ||
DB 577 GCGTCTCTGAGGTGCACACACCGGCTTACGCCGAGAGACGCGCGGCGGACGCTCG 518

QY 302 atcgtcgttgattatgattgaaatgagaagtgtcttaattggaaggaataaccg 361
    || || || || || || || || || || || || || || || || || || || ||
DB 517 ACCAGGCGGCGGACGACACGATCGGCCCTCAAGACGTCCTCGCGCGGACGAGACTRACG 458

QY 362 tcaataataatgtagaataatggttctgttgtaggaaggagatgcttggtgacg 421
    || || || || || || || || || || || || || || || || || || || ||
DB 457 TCATCTGTCACATCGGCAACGAGCCCTGGGGCAACACCGACCCCGCGCTGCACGAGC 398

```

```

OY 422 ggtataaacaagcaatcccgagattgctaaagccggtctaaccataccttgatgtag 481
DB 397 CGACCGTCGCGCGCCCTCAAGAGCTCGCGCGCGGCTCCAGCACAGATCAAGTG 338
OY 482 atgtctgaggggtgggaattcccaatcg---atcatattatgagaagagttt 538
DB 337 ACGCCCCCACTGGGGCCAGACTGCGAGGGCGTATGCTGCCAAGCAGCGTCCGCT 278
OY 539 ttaatgctgacctcaacgaataacatgatttcgattcatatgta 584
DB 277 ACGAGCGCGACCCCAACGCAACTGATCTCTCATCCACATGTA 232

RESULT 11
AC024591 111071 bp DNA HTG 18-JUL-2000
LOCUS Homo sapiens chromosome 16 clone RP11-511G21, WORKING DRAFT
DEFINITION
SEQUENCE, 13 ordered pieces.
AC024591
AC024591.3 GI:9256450
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 111071)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
ENTRY 2 (bases 1 to 111071)
DOE Joint Genome Institute.
DIRECT SUBMISSION
Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7705016.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 602754
Center clone name: RPCT-11_511G21
-----
Summary Statistics
Consensus quality: 104021 bases at least Q40
Consensus quality: 108348 bases at least Q30
Consensus quality: 109319 bases at least Q20
Estimated insert size: 118930; agarose-1p estimation
Estimated insert size: 110521; sum-of-ctotigs estimation
Quality coverage: 8.44 in Q20 bases; agarose-1p estimation
Quality coverage: 9.08 in Q20 bases; sum-of-ctotigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
* 3999: contig of 3999 bp in length
* 4000 4099: gap of unknown length
* 4100 6224: contig of 2125 bp in length
* 6225 6324: gap of unknown length
* 6325 14602: contig of 8278 bp in length
* 14603 14702: gap of unknown length
* 14703 34245: contig of 39543 bp in length
* 34246 54345: gap of unknown length
* 54346 58569: contig of 4224 bp in length
* 58570 58670: gap of unknown length
* 58670 59480: contig of 811 bp in length

```

```

* 59481 59580: gap of unknown length
* 59581 60683: contig of 1103 bp in length
* 60684 60783: gap of unknown length
* 60784 62950: contig of 2167 bp in length
* 62951 63050: gap of unknown length
* 63051 64092: contig of 1042 bp in length
* 64093 64192: gap of unknown length
* 64193 85067: contig of 20875 bp in length
* 85068 85167: gap of unknown length
* 85168 91249: contig of 5982 bp in length
* 91250 108236: contig of 16987 bp in length
* 108237 108335: gap of unknown length
* 108337 111071: contig of 2735 bp in length.
Location/Qualifiers
source 1..111071
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-511G21"
/clone_id="RPCT human BAC library 11"
BASE COUNT 27517 a 27152 c 28256 g 26944 t 1202 others
ORIGIN
Query Match 5.18; Score 45.8; DB 53; Length 111071;
Best Local Similarity 34.7%; Pred. No. 0.6;
Matches 110; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
OY 232 gaagataatcattggtgctgcttgaagttatgtagtgcacggatgattccatt 291
DB 59292 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59351
OY 292 gcttcgcccaatgctgctgcttgaattgattgattgaaagaaagtcttaataag 351
DB 59352 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59411
OY 352 gaagataatcattgattgattgattgattgattgattgattgattgattgatt 411
DB 59412 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59471
OY 412 tgggtgacgggtataaacaagcaatcccgagattgctaaagccggtctaaccatcc 471
DB 59472 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59531
OY 472 ttgatgtagatgctgctgctgctgctgctgctgctgctgctgctgctgctgct 531
DB 59532 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 59591
OY 532 gaagtttttaattgctga 548
DB 59592 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 59608

RESULT 12
CELK03E6
LOCUS CELK03E6 35049 bp DNA INV 28-OCT-1998
DEFINITION Caenorhabditis elegans cosmid K03E6.
ACCESSION U55375
VERSION U55375.1 GI:3805671
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditidae; Rhabditiidae; Poloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 35049)
AUTHORS Wilson,R., Ahnscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,U.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurry,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Koopra,A.,

```

saunders,D., Shownkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
 Staaden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
 Vaughan,K., Waterston,R., Watson,A., Wernstock,L.,
 Wilkinson-Spoot,J. and Woldman,P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*
 Nature 368 (6466), 32-38 (1994)
 JOURNAL MEDLINE 94150718
 REFERENCE 2 (bases 1 to 35049)
 AUTHORS Latreille,P. and Gattung,S.
 TITLE The sequence of *C. elegans* cosmid K03E6
 JOURNAL Unpublished (1998)
 REFERENCE 3 (bases 1 to 35049)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (17-APR-1996) Robert Waterston
 REFERENCE 4 (bases 1 to 35049)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (21-APR-1997)
 REFERENCE 5 (bases 1 to 35049)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (28-OCT-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Oct 28, 1998 this sequence version replaced g1:1280123.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 e-mail: twenematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is C04D1, 2700 bp overlap; 3' clone is F49E7, 200 bp
 overlap. Actual start of this clone is at base position 701 of
 CELK03E6; actual end is at 35049 of CELK03E6

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program Genefinder(P. Green and L. Hillier, ms in preparation).
 Location/Qualifiers

FEATURES
 source
 1. 35049
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="K03E6"
 /chromosome="X"
 complement(3691..7748)
 /gene="unc-1"
 complement(join(3691..3766,3931..4145,4822..5095,
 5154..5278,6432..6545,7296..7326,7735..7748))
 /gene="unc-1"
 /note="K03E6.5"
 /codon_start=1
 /product="Erythrocyte band 7 intergal membrane protein"

gene
 CDS
 complement(join(12649..14893)
 /gene="K03E6.4"
 complement(join(12649..12750,12801..12928,13397..13527,
 13570..13666,14217..14391,14837..14893))
 /gene="K03E6.4"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAC69043.1"
 /db_xref="GI:1280126"

gene
 CDS
 complement(join(20796..20881,21727..21835,22141..22262,23469..23598,
 23885..24151,24209..24499)
 /gene="11m-6"
 /note="K03E6.1"
 /codon_start=1
 /product="similar to other homeobox domains; contains two
 LIM domains"
 /protein_id="AAC69042.1"
 /db_xref="GI:1280127"

gene
 CDS
 complement(join(25406..26485)
 /gene="K03E6.6"
 complement(join(25406..25503,25556..25679,25726..25773,
 26375..26485))
 /gene="K03E6.6"
 /note="similar to Profilin; K03E6.6"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAC69045.1"
 /db_xref="GI:1280128"

gene
 CDS
 complement(join(28061..28224,28960..29077,30844..30997,
 31048..31205,31305..31639,32214..32247,33211..33354,
 33555..33654,33708..33861,33914..33962))
 /gene="K03E6.7"
 /note="similarity to *C. elegans* protein C01C10.4"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAC69046.1"
 /db_xref="GI:1280129"

gene
 CDS
 complement(join(28061..28224,28960..29077,30844..30997,
 31048..31205,31305..31639,32214..32247,33211..33354,
 33555..33654,33708..33861,33914..33962))
 /gene="K03E6.7"
 /note="similarity to *C. elegans* protein C01C10.4"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAC69046.1"
 /db_xref="GI:1280129"

Oy 638 ttaatcaagacctcgcatatgtagtgaatttgacacccgtcatatacaaatgtgacg 697
Db 19456 CCAATCAAGATCTCAGCTGGCTTTTGTGAGGAACTGGGCGCTGTGATTCTAAATTATTA 19397
Oy 698 tcgatgaagcaacgattatgag 719
Db 19396 TAGAAATGCAAAAGACCCAGAG 19375

Search completed: December 20, 2000, 03:17:03
Job time: 13068 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:58:32 ; Search time 168.09 Seconds
(without alignments)
2011.401 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990
Perfect score: 900
Sequence: 1 aatgcacatccgcgatttta.....gattaaacacacgttttaca 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- 1: N_Geneseq_36.*
- 2: /SIDS6/gcgdata/geneseq/geneseqn/NA1980.DAT.*
- 3: /SIDS6/gcgdata/geneseq/geneseqn/NA1981.DAT.*
- 4: /SIDS6/gcgdata/geneseq/geneseqn/NA1982.DAT.*
- 5: /SIDS6/gcgdata/geneseq/geneseqn/NA1983.DAT.*
- 6: /SIDS6/gcgdata/geneseq/geneseqn/NA1984.DAT.*
- 7: /SIDS6/gcgdata/geneseq/geneseqn/NA1985.DAT.*
- 8: /SIDS6/gcgdata/geneseq/geneseqn/NA1986.DAT.*
- 9: /SIDS6/gcgdata/geneseq/geneseqn/NA1987.DAT.*
- 10: /SIDS6/gcgdata/geneseq/geneseqn/NA1988.DAT.*
- 11: /SIDS6/gcgdata/geneseq/geneseqn/NA1989.DAT.*
- 12: /SIDS6/gcgdata/geneseq/geneseqn/NA1990.DAT.*
- 13: /SIDS6/gcgdata/geneseq/geneseqn/NA1991.DAT.*
- 14: /SIDS6/gcgdata/geneseq/geneseqn/NA1992.DAT.*
- 15: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT.*
- 16: /SIDS6/gcgdata/geneseq/geneseqn/NA1994.DAT.*
- 17: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT.*
- 18: /SIDS6/gcgdata/geneseq/geneseqn/NA1996.DAT.*
- 19: /SIDS6/gcgdata/geneseq/geneseqn/NA1997.DAT.*
- 20: /SIDS6/gcgdata/geneseq/geneseqn/NA1998.DAT.*
- 21: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.*
- 22: /SIDS6/gcgdata/geneseq/geneseqn/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 900 | 100.0 | 1470 | 21 | DNA encoding a Bac |
| 2 | 895.4 | 99.5 | 1438 | 21 | DNA encoding a man |
| 3 | 468.2 | 52.0 | 1407 | 21 | Bacillus agaradher |
| 4 | 468.2 | 52.0 | 1407 | 21 | DNA encoding a Bac |
| 5 | 468.2 | 52.0 | 1482 | 21 | Bacillus agaradher |
| 6 | 468.2 | 52.0 | 1482 | 21 | DNA encoding a Bac |
| 7 | 408.4 | 45.0 | 1107 | 21 | DNA encoding a Bac |
| 8 | 405 | 45.0 | 995 | 21 | DNA encoding a Bac |
| 9 | 304.4 | 33.8 | 960 | 21 | DNA encoding a Bac |
| 10 | 157.8 | 18.6 | 564 | 21 | DNA encoding a par |
| 11 | 104.4 | 11.6 | 397 | 21 | DNA encoding a par |
| 12 | 78.2 | 8.7 | 915 | 21 | DNA encoding a par |

| | | | | | |
|----|------|-----|--------|----|--------|
| 13 | 35 | 3.9 | 2040 | 17 | T29774 |
| 14 | 35 | 3.9 | 2100 | 10 | M90712 |
| 15 | 35 | 3.9 | 2901 | 12 | Q14810 |
| 16 | 35 | 3.9 | 3558 | 13 | Q22596 |
| 17 | 34.6 | 3.8 | 580073 | 18 | T58840 |
| 18 | 33.6 | 3.7 | 627 | 20 | Z08233 |
| 19 | 33.6 | 3.7 | 1395 | 20 | Z08233 |
| 20 | 33.6 | 3.7 | 1680 | 20 | Z08234 |
| 21 | 33.6 | 3.7 | 1763 | 15 | O55316 |
| 22 | 33.4 | 3.7 | 2408 | 12 | V17621 |
| 23 | 33.4 | 3.7 | 2802 | 19 | O14809 |
| 24 | 33.2 | 3.7 | 633 | 18 | V75647 |
| 25 | 33.2 | 3.7 | 1800 | 20 | Z25012 |
| 26 | 32.6 | 3.6 | 138169 | 21 | A34791 |
| 27 | 32.6 | 3.6 | 141589 | 21 | A35005 |
| 28 | 32.6 | 3.6 | 141589 | 21 | A35030 |
| 29 | 32.4 | 3.6 | 1300 | 20 | V82078 |
| 30 | 32.4 | 3.6 | 2132 | 20 | V82077 |
| 31 | 32.4 | 3.6 | 2841 | 18 | V74488 |
| 32 | 32.4 | 3.6 | 3006 | 21 | Z29484 |
| 33 | 32.4 | 3.6 | 22080 | 20 | X06751 |
| 34 | 32.2 | 3.6 | 1021 | 16 | T04613 |
| 35 | 32.2 | 3.6 | 1026 | 17 | T13340 |
| 36 | 32.2 | 3.6 | 1399 | 20 | X20017 |
| 37 | 32.2 | 3.6 | 1500 | 20 | X20016 |
| 38 | 32.2 | 3.6 | 3111 | 16 | O87260 |
| 39 | 32.2 | 3.6 | 4956 | 20 | X13728 |
| 40 | 32.2 | 3.6 | 1448 | 18 | X14454 |
| 41 | 32 | 3.6 | 4359 | 18 | T69969 |
| 42 | 32 | 3.6 | 4365 | 14 | O52440 |
| 43 | 32 | 3.6 | 4365 | 13 | O25536 |
| 44 | 32 | 3.6 | 4365 | 13 | O25534 |
| 45 | 32 | 3.6 | 4365 | 14 | O52443 |

ALIGNMENTS

| | |
|----------------|--------------------------------------------------------------------------|
| RESULT 1 | |
| ID 245335 | standard; DNA; 1470 BP. |
| XX 245335; | |
| AC | |
| XX | |
| DT 27-MAR-2000 | (first entry) |
| XX | |
| DE | DNA encoding a Bacillus mannanase enzyme. |
| XX | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; |
| KW | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; |
| KW | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; |
| KW | printing paste; plant material degradation; recycled waste paper; |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; |
| KW | mannan-containing food; coffee extract; cleaning composition; |
| KW | machine washing; hard surface cleaner; dishwashing; oral; dental; |
| KW | contact lens; body care composition; fabric softener; oil well drilling; |
| KW | subterranean formation fracture; ss. |
| XX | |
| OS | Bacillus sp. |
| XX | |
| EH | |
| FT | Location/Qualifiers |
| FT | 1..1470 |
| FT | 1..1470 |
| FT | /*tag- a |
| FT | /product- "mannanase" |
| FT | /transl_except- (pos: 346..348, aa: Pro) |
| FT | /note- "no termination codon given" |
| FT | sig_peptide |
| FT | 1..90 |
| FT | /*tag- b |
| FT | 91..1470 |
| FT | mat_peptide |
| FT | 1..1470 |
| FT | /*tag- c |
| FT | /*tag- c |
| FT | "specifically claimed in claim 4; nucleotides |
| FT | 91-990 encode the catalytically active core" |


```

XX  MO9964619-A2.
XX  16-DEC-1999.
XX  10-JUN-1999; 99WO-DK00314.
XX  10-JUN-1998; 98US-0111256.
PR  20-OCT-1998; 98DK-0001340.
PR  20-OCT-1998; 98DK-0001341.
PR  28-OCT-1998; 98US-0105970.
PR  28-OCT-1998; 98US-0106054.
PR  23-DEC-1998; 98DK-0001725.
PR  05-MAR-1999; 99DK-0000306.
PR  05-MAR-1999; 99DK-0000307.
PR  05-MAR-1999; 99DK-0000308.
PR  05-MAR-1999; 99DK-0000309.
PR  09-MAR-1999; 99US-0123543.
PR  10-MAR-1999; 99US-0123623.
PR  10-MAR-1999; 99US-0123641.
PR  11-MAR-1999; 99US-0123642.
XX  (NOVO ) NOVO-NORDISK AS.
XX  Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX  WPI: 2000-105891/09.
XX  P-PSDB: Y54122.
XX  New mannanses for treatment of textiles, plant material and coffee
XX  extract, and in cleaning compositions
XX  Claim 4; Page 208; 242pp; English.
XX  The present sequence encodes a mannanase (also known as mannan
XX  endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
XX  of Bacillus sp. 1633. The mannanase hydrolyses galactomannans.
XX  Specifically, mannanses hydrolyse 1,4'-beta-D-mannosidic linkages in
XX  mannans, galactomannans, glucomannans, and galactoglucomannans. The
XX  mannanase protein, or preparations containing it, are used to improve
XX  properties of cellulosic or synthetic fibres, yarn or (non)woven
XX  fabrics (removal of mannan-based sizes or printing pastes). They are
XX  also used to degrade or modify plant materials (particularly recycled
XX  waste paper, paper making pulps, or material containing guar or locust
XX  bean gums (thickeners), or to reduce viscosity of mannan-containing
XX  foods or feeds). The mannanses are also used to process coffee
XX  extracts (to inhibit gel formation); in cleaning compositions (for
XX  machine washing of fabrics, as hard-surface cleaners, for hand or
XX  machine dishwashing, also in oral, dental, contact lens or body-care
XX  compositions) where they remove mannan-containing soils and prevent
XX  binding of some soils to cellulosics; and in fabric softeners. They
XX  can also be used in oil well drilling to fracture subterranean
XX  formations.
XX  Sequence 1470 BP; 441 A; 237 C; 355 G; 437 T; 0 other:

```

Query Match 100.0%; Score 900; DB 21; Length 1470;
 Best Local Similarity 100.0%; Pred. No. 1.4e-261;
 Matches 900; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 aatgaacattccggaatttatgttaagcgggtacacattctatacgaatgcaatgaaacca 60
    |||||
DB 91 aatgaacattccggaatttatgttaagcgggtacacattctatacgaatgcaatgaaacca 150
OY 61 ttgtgaagagaggggattacacattggcgcgcatggtataaagaccaggaactactcga 120
    |||||
DB 131 ttgtgaagagaggggattacacattggcgcgcatggtataaagaccaggaactactcga 210
OY 121 attgaagagattgcaataacacggtgctatacgaatggtgtttctcgaatgggga 180
    |||||
DB 211 attgaagagattgcaataacacggtgctatacgaatggtgtttctcgaatgggga 270

```

```

OY 181 caatgacaaaagatgacatccatagaagaacattatctctttagcgggaataat 240
    |||||
DB 271 caatgacaaaagatgacatccatagaagaacattatctctttagcgggaataat 330
OY 241 cattgtgtgtctgtcttgaaagtatcatgatgtacacggttatgatcttcgttcgc 300
    |||||
DB 331 cattgtgtgtctgtcttgaaagtatcatgatgtacacggttatgatcttcgttcgc 390
OY 301 aatcgtcgtgtgtatattgattgaaatggaagtgcttaatttgaaagaagatacc 360
    |||||
DB 391 aatcgtcgtgtgtatattgattgaaatggaagtgcttaatttgaaagaagatacc 450
OY 361 gcatattataattgcgaatgaaatggtgttgctgttggaagggagatgttggttac 420
    |||||
DB 451 gcatattataattgcgaatgaaatggtgttgctgttggaagggagatgttggttac 510
OY 421 ggtataaacaagcaatcccgatgtgtaacgcggtctcaaacacattcttgta 480
    |||||
DB 511 ggtataaacaagcaatcccgatgtgtaacgcggtctcaaacacattcttgta 570
OY 481 gatgtcgggggtgvggacaattccacaatcgattcatatgatabgaagaagtttt 540
    |||||
DB 571 gatgtcgggggtgvggacaattccacaatcgattcatatgatabgaagaagtttt 630
OY 541 aatgtgcacctcagcaataataatgtttcgattcatatgatabgaagaagttgt 600
    |||||
DB 631 aatgtgcacctcagcaataataatgtttcgattcatatgatabgaagaagttgt 690
OY 601 aatgcatcgaagtctgactaataattgaccgagttcttaatacgaactcgtcatatgc 660
    |||||
DB 691 aatgcatcgaagtctgactaataattgaccgagttcttaatacgaactcgtcatatgc 750
OY 661 attgtgaatttggaacacgtctacatacaaatgtgtacgtcgtatgaaagcaatgagc 720
    |||||
DB 751 attgtgaatttggaacacgtctacatacaaatgtgtacgtcgtatgaaagcaatgagc 810
OY 721 tatcttgacaagaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
    |||||
DB 811 tatcttgacaagaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 870
OY 781 tgggaattatgaacatttcgaatgattggcttggaataaaccttacagcttgvggaat 840
    |||||
DB 871 tgggaattatgaacatttcgaatgattggcttggaataaaccttacagcttgvggaat 930
OY 841 acaatgagatgtgtcattatggttttaagagaacattcgagattaaagcagcttttaca 900
    |||||
DB 931 acaatgagatgtgtcattatggttttaagagaacattcgagattaaagcagcttttaca 990
    |||||

```

RESULT 2
 245336
 ID 245336 standard; DNA; 1438 BP.
 XX 245336;
 DT 27-MAR-2000 (first entry)
 XX
 DE DNA encoding a mannanase-linker-cellulose binding domain fusion.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
 KW 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
 KW galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
 KW printing paste; plant material degradation; recycled waste paper;
 KW paper making pulp; guar; locust bean gum; thickener; viscosity;
 KW mannan-containing food; coffee extract; cleaning composition;
 KW machine washing; hard-surface cleaner; dishwashing; oral; dental;
 KW contact lens; body-care composition; fabric softener; oil well drilling;
 KW subterranean formation fracture; cellulose binding domain; ss.
 XX
 OS Synthetic.
 OS Bacillus sp.
 OS Clostridium thermocellum.

```

XX FH Key Location/Qualifiers
XX FT CDS 1..1431
XX FT /*tag- a
XX PN MO9964619-A2.
XX PD 16-DEC-1999.
XX PF 10-JUN-1999; 99WO-DK00314.
XX PR 10-JUN-1998; 98US-0111256.
XX PR 20-OCT-1998; 98DK-0001340.
XX PR 20-OCT-1998; 98DK-0001341.
XX PR 28-OCT-1998; 98US-0105970.
XX PR 28-OCT-1998; 98US-0106054.
XX PR 23-DEC-1998; 98DK-0001725.
XX PR 05-MAR-1999; 99DK-0000306.
XX PR 05-MAR-1999; 99DK-0000307.
XX PR 05-MAR-1999; 99DK-0000308.
XX PR 05-MAR-1999; 99DK-0000309.
XX PR 09-MAR-1999; 99US-0123543.
XX PR 10-MAR-1999; 99US-0123623.
XX PR 10-MAR-1999; 99US-0123641.
XX PR 11-MAR-1999; 99US-0123642.
XX PA (NOVO ) NOVO-NORDISK AS.
XX PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX DR WPI: 2000-105891/09.
XX DR P-PSDB: Y54123.
XX PT New mannases for treatment of textiles, plant material and coffee
XX PT extract, and in cleaning compositions -
XX PS Example 4; Page 210; 242pp; English.
XX PA
XX CC The present sequence encodes a mannase-linker-cellulose binding
XX CC domain fusion protein. Mannase (also known as mannan
XX CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
XX CC hydrolyses galactomannans. Specifically, mannases hydrolyse
XX CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
XX CC and galactoglucomannans. The mannase protein, or preparations
XX CC containing it, are used to improve properties of cellululosic or
XX CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX CC sites or printing pastes). They are also used to degrade or modify
XX CC plant materials (particularly recycled waste paper, paper making pulps,
XX CC or material containing guar or locust bean gums (thickeners), or to
XX CC reduce viscosity of mannan-containing foods or feeds). The mannases
XX CC are also used to process coffee extracts (to inhibit gel formation);
XX CC in cleaning compositions (for machine washing of fabrics, as
XX CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX CC dental, contact lens or body-care compositions) where they remove
XX CC mannan-containing soils and prevent binding of some soils to
XX CC celluloses; and in fabric softeners. They can also be used in oil
XX CC well drilling to fracture subterranean formations.
XX SO Sequence 1438 BP; 445 A; 267 C; 344 G; 382 T; 0 other;

Query Match 99.5%; Score 895.4; DB 21; Length 1438;
Best Local Similarity 99.9%; Pred. No. 3.4e-260;
Matches 896; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 gcaattcggattttagtaagcggtaccactctatcgatgccaatggaaacccattt 63
DB 1 gcaattcggattttagtaagcggtaccactctatcgatgccaatggaaacccattt 60
OY 64 gtaatgagaggatttaacatcgagcagcatggtataagaccaggaactactgcaatt 123
DB 61 gtaatgagaggatttaacatcgagcagcatggtataagaccaggaactactgcaatt 120

```

```

OY 124 gaaggattgcaaaaccgggtgctaatcggtccggattggttctcctgatggggacaa 163
DB 121 gaaggattgcaaaaccgggtgctaatcggtccggattggttctcctgatggggacaa 180
OY 184 tggacaagaatgataccatcaacagtaagaaccttactctcttagcggagatatcat 243
DB 181 tggacaagaatgataccatcaacagtaagaaccttactctcttagcggagatatcat 240
OY 244 ttggttgcgtcttcttgaaagttcatgctacacgggttagtattccattgcttcgcat 303
DB 241 ttggttgcgtcttcttgaaagttcatgctacacgggttagtattccattgcttcgcat 300
OY 304 cgtgcgttgcattatttgatggaatggaagctcttaatttggaagaagatccgtc 363
DB 301 cgtgcgttgcattatttgatggaatggaagctcttaatttggaagaagatccgtc 360
OY 364 attattaatttgcaagaatggttgcgttcgttggaaggaggtcgttgcaagg 423
DB 361 attattaatttgcaagaatggttgcgttcgttggaaggaggtcgttgcaagg 420
OY 424 tataacaagaacatcccgatgctgtaacgcgggtctaaacatccttgatgtagat 483
DB 421 tataacaagaacatcccgatgctgtaacgcgggtctaaacatccttgatgtagat 480
OY 484 gctgcgggttgaggacaattccacaatcgattcatgattatggaaggagattttat 543
DB 481 gctgcgggttgaggacaattccacaatcgattcatgattatggaaggagattttat 540
OY 544 gctgacctcaacgaataacatgcttccattcatgattatggaaggagattttat 603
DB 541 gctgacctcaacgaataacatgcttccattcatgattatggaaggagattttat 600
OY 604 gcatcgcaagtcgtactaatattgacaggagttcttaatacgaacctgcattagcat 663
DB 601 gcatcgcaagtcgtactaatattgacaggagttcttaatacgaacctgcattagcat 660
OY 664 ggtgaatttgacacgcgtcatcaaaaatggtgacgtcgatgaagaagcaagattagctat 723
DB 661 ggtgaatttgacacgcgtcatcaaaaatggtgacgtcgatgaagaagcaagattagctat 720
OY 724 tctgacaagaaggaggttgggtggttgcgttcgtgcataaggaaaggagcccaagatgg 783
DB 721 tctgacaagaaggaggttgggtggttgcgttcgtgcataaggaaaggagcccaagatgg 780
OY 784 gagtattgaccttcgcaatgattggcttggaataacattacagcttggggaataca 843
DB 781 gagtattgaccttcgcaatgattggcttggaataacattacagcttggggaataca 840
OY 844 atagtgaatgctcatatggtttaagagaacctcgagattaaagcagcgtttttaca 900
DB 841 atagtgaatgctcatatggtttaagagaacctcgagattaaagcagcgtttttaca 897

RESULT 3
Z29846
ID Z29846 standard; DNA; 1407 BP.
AC Z29846;
XX 27-MAR-2000 (first entry)
DE Bacillus agaradherens Clone MB594, Mannanase encoding DNA.
XX Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing;
XX cosmetic stain; food stain; ss.
XX Bacillus agaradherens.
OS Synthetic.
XX Key Location/Qualifiers
XX FT 1..1407
XX FT /*tag- a

```

```

FT      /product- "Mannanase"
FT      /EC_number- "3.2.1.78"
FT      /function- "Hydrolysis of 1,4-beta-D-mannosidic linkages"
FT      sig_peptide      1..93
FT      mat_peptide      94..1404
FT      /tag- b
FT      /tag- C
FT      /label- Mature_Mannanase
XX      W09964552-A1.
XX      16-DEC-1999.
XX      PD
XX      PF 10-JUN-1998; 98MO-US12026.
XX      PR 10-JUN-1998; 98MO-US12026.
XX      PA (PROC ) PROCTER & GAMBLE CO.
XX      PI Bettiol JP, Thoen CAJK;
XX      DR WPI: 2000-116536/10.
XX      DR P-PSDB: Y44496.
XX      PT Detergent composition for removing greasy stains such as cosmetics,
XX      PT food stains and body soils
XX      PS Disclosure: Page 102-103; 113pp; English.
XX      CC The present sequence is the B. agaradherens Clone MB594 DNA encoding
XX      CC alkaline mannanase enzyme. It is derived from B. agaradherens strain
XX      CC NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
XX      CC from 7.5-10.5. It can be used in a detergent composition along with a
XX      CC mid-branched anionic surfactant. The detergent composition may be used
XX      CC for washing purposes, soaking/pre-treatment of stained fabric, hard
XX      CC surface cleaning and for removal of cosmetic and/or food stains. This
XX      CC composition provides excellent cleaning effect at low temperature.
XX      SQ Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other:

```

```

Query Match      52.0%; Score 46812; DB 21; Length 1407;
Best Local Similarity 70.1%; Pred. No. 2e-131;
Matches 629; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

```

```

OY      4 gcaaatccggaattatgtaagcggtaccactctacgacgtgcaatggaaccattt 63
DB      94 gcaagttacagcgctttatgtgtgcaatacgttatatgacgcaaatgsgcagccatt 153
OY      64 gtaatgaggggattaaacatgagcgatggtatataagacagcgcaactactgcaatt 123
DB      154 gtcagtaggggtattacacatgacatgcttgttataagaacacgctccaacagttatt 213
OY      124 gaaggagattgcaaatccggtgctaatacgtgcggatgtgttattcgtatgsggagcaa 183
DB      214 cctgcattgcagaggaagaaacgcaacgattcgattgtttatcagatggcggtcaa 273
OY      184 tggacaagaagatgacatcatcactgaagaacacttactctttagcggaataatctt 243
DB      274 tgggaaaagaagacacatgacacacatctgtaagtcattgagcatctgaggaacaataaa 333
OY      244 ttggttcgcttctgaagtattatgattacggttattccattccttcgctgcaatt 303
DB      334 atggtagcgtgctgttgaagttacatgacgaggtcgcgattcgcgcagtttaatt 393
OY      304 cgtgctgtgattattggaattgaatgagaagtgctttaattggaagaagaataccgtc 363
DB      394 cgaagcgttattgatttggtagaagaatgaaagattcgcttaataagaagatacgtt 453
OY      364 attataatatgcaatgaaatggtttggttcgttgggaagggagatgcttggtcgaagg 423
DB      454 attataacattgcaaaacgaagtgataggaattggagtgctcagcttggtgcagattgac 513

```

```

OY      424 tataacaagaatcccgcgattgcgtacgacggttcaaacctacttattgtagat 483
DB      514 tatattgtatcattccgaagcttcgcgtatcgcgcttaaacacacacacttaattgtagat 573
OY      484 gctgcggggtgggacattcccaacatgataatgataatggaagaagattttat 543
DB      574 gcaagcagatggtgggcaataatcccgaaactctatcattcgtttcggacaagaatgtagat 633
OY      544 gctgacctcaacgaataacataatggttttcgattcattatgataatgacagtgtagat 603
DB      634 gcaatccgttaaaaaatacagatgcttcacatcattatgatatgtagtgcgtgtagat 693
OY      604 gcatcgcaagttcgtactcaatattgacaggaattcttcaataagacctgacttagtact 663
DB      694 gctaacactgttagatccaatattgtatagatcattagatcagaacctgcttcctgtaata 753
OY      664 gttgaatttgacacacgctacataaatgtgtacgtcgtatgaagaacagattatgtagat 723
DB      754 gttgaattcgttcataagacatacatgattgtgtatgtatgaagaatacactccttagtact 813
OY      724 tctgacaagaagagagttggtggtgtgctgtgtcatatgaaaggaagcagccacgaatg 783
DB      814 tctgagaagaactgacacaggtgtgctgcgttctgtcttgaaaggaacagatcaccgaatg 873
OY      784 gagatttagaccttcgaatgattggtgctggaataacacttaacgcttggtgggaataga 843
DB      874 gactatttagaccttcgaagacgtggtcgtgtaacacttaacgatttggtgggaataga 933
OY      844 atagtgaatgtccatattgattgaagaagaactcgaagattgaacgacgcttttaca 900
DB      934 attgcccaggggcccagatggtgttaacgaagaacccccaacatccacgctatttaca 990

```

```

RESULT 4
245338
245338 standard; DNA; 1407 BP.
XX
XX 245338;

```

```

DT 27-MAR-2000 (first entry)
XX
XX

```

```

DE DNA encoding a Bacillus mannanase enzyme.
XX
XX

```

```

KM Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
OS mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
XX synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KM plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
XX cleaning composition; machine washing; hard-surface cleaner;
KM dishwashing; oral; dental; contact lens; body-care composition;
XX fabric softener; oil well drilling; subterranean formation fracture; ss.

```

```

OS Synthetic.
XX Bacillus agaradherens.
XX
XX

```

```

FE Key location/Qualifiers
FT CDS 1..1407
FT /tag- a
FT /product- "mannanase"

```

```

XX      W09964619-A2.
XX      16-DEC-1999.
XX      PD
XX      PF 10-JUN-1999; 99MO-DK00314.
XX      XX
XX      PR 10-JUN-1998; 98US-0111256.
XX      PR 20-OCT-1998; 98DK-0001340.
XX      PR 20-OCT-1998; 98DK-0001341.
XX      PR 28-OCT-1998; 98US-0105970.
XX      PR 28-OCT-1998; 98US-0106054.
XX      PR 23-DEC-1998; 98DK-0001725.

```

PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX WPI: 2000-105891/09.
 DR P-PSDB: 154125.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 XX extract, and in cleaning compositions -
 XX
 XX Example 5; Page 214-215; 242pp; English.
 XX
 XX The present sequence encodes a *Bacillus mannanase* (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The present mannase is a synthetic variant of the mannase of
 CC Y54124, in which the C-terminus of the protein was changed due to
 CC design of a lower PCR primer used for amplification. The mannase
 CC hydrolyses galactomannans. Specifically, mannases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC slimes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 CC
 CC Sequence 1407 BP; 424 A; 272 C; 334 G; 377 T; 0 other:
 XX

Query Match 52.0%; Score 468.2; DB 21; Length 1407;
 Best Local Similarity 70.1%; Pred. NO. 2e-131;
 Matches 629; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

QY 364 attataatattcgcaatgaaatggttgcgttcgtggaaggaggtgcttgagctacg 423
 DB 454 attataatattcgcaatgaaatggttgcgttcgtggaaggaggtgcttgagctacg 513
 QY 424 tataaacaagaatcccgatgctgtaacggcgttcaaacacacttggttgat 483
 DB 514 tataatgctatcccgaaagcttcgagcttcgaacacacacttggttgat 573
 QY 484 gctgggggtgggacaattccacaatcgatctgataatgaaagaagtattat 543
 DB 574 gcagcagatgggggcataatccgcaatctatctatgattgaagcaagatggttat 633
 QY 544 gctgacctcaagcaatacaatgtttcgcattcatatgatatgaaatgagtgat 603
 DB 634 gcagatccgttaaaaaatacagatgtctccatccatcatgatagtatgctgtggtat 693
 QY 604 gcatcgcaagtgcgtacaaatattgacagagttcttaacagacctgcattgctat 663
 DB 694 gctaactgttaagatacaataatgataagcatagatcaagacctgctctcgtata 753
 QY 664 gttgaattggaacacgcatacaaatgtgacgtcgatgaagcaacgattatgacctat 723
 DB 754 gttgattcgcgcataagacatactatgattgattgattgaagatacaatccttgat 813
 QY 724 tctgaacaagaagaggtggtgtgctggtgctatgaaaggaaggaacgcccagaatg 783
 DB 814 tctgaagaagaatgacacaggtggtgctgctgttcttgaaaggaacagatccgactg 873
 QY 784 gagtttagaccttgaaatgattgaggtgctgaaataacattacagcttgaggaaatga 843
 DB 874 gactattagaccttcgaagaagctggtgctgaacattacagcttgaggaaatga 933
 QY 844 atagtaagtgtcatatggtttaagaagaactcgaaatgaatgaacagcttttaca 900
 DB 934 attgccaagggtgcagatggtcttaagagaacacctcaacatccacgctatttaca 990

RESULT 5
 229845
 ID 229845 standard; DNA; 1482 BP.
 XX
 AC 229845;
 DT 27-MAR-2000 (first entry)
 DE *Bacillus agaradherens* NCIMB 40482, Mannanase encoding DNA.
 XX
 KW Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
 KW detergent composition; mid-branched anionic surfactant; washing;
 KW cosmetic stain; food stain; ss.
 XX
 OS *Bacillus agaradherens*.
 XX
 FH key Location/Qualifiers
 FT CDS 1..1482
 FT /tag a
 FT /product "Mannanase"
 FT /EC number "3.2.1.78"
 FT /function "Hydrolysis of 1,4-beta-D-mannosidic linkages"
 FT sig_peptide 1..96
 FT /tag b
 FT mat_peptide 97..1029
 FT /tag c
 FT /label Mature_Mannanase
 XX
 FN W09964552-A1.
 XX
 PD 16-DEC-1999.
 XX
 PD 10-JUN-1998; 98WO-US12026.
 XX
 PD 10-JUN-1998; 98WO-US12026.
 PR

PA (PROC) PROCTER & GAMBLE CO.
XX
XX Bettiol JP, Thoen CAJK;
PI
XX WPI: 2000-116536/10.
DR
DR P-PSDB: Y44495.
XX
XX Detergent composition for removing greasy stains such as cosmetics,
PT food stains and body soils
PT
XX
XX Disclosure; Page 101; 113pp; English.
PS
XX
XX The present sequence is a DNA encoding B. agaradherens NCIMB 40482,
CC alkaline mannanase enzyme. Mannanase shows maximum activity at pH ranging
CC from 7.5-10.5. It can be used in a detergent composition along with a
CC mild-branched anionic surfactant. The detergent composition may be used
CC for washing purposes, soaking/pre-treatment of stained fabric, hard
CC surface cleaning and for removal of cosmetic and/or food stains. This
CC composition provides excellent cleaning effect at low temperature.
XX
XX Sequence 1482 BP; 446 A; 285 C; 352 G; 339 T; 0 other;

| | | | | |
|-----------------------|-----------------|---------------------|-----------|--------------|
| Query Match | 52.0% | Score 46812; | DB 21; | Length 1482; |
| Best Local Similarity | 70.1%; | Pred. No. 2.1e-131; | | |
| Matches 629; | Conservative 0; | Mismatches 268; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|-----------------------------------------------------------------------|-----|
| QY | 4 | gcaaatccggatatttattgtaaaagcgtacacccctctatagaagtgcgaattggaacccattt | 63 |
| Db | 94 | gcaaatccggatcggcttttatgtctatgtaatgcataatcgttatatgacgaataatgvcagccattc | 153 |
| QY | 64 | gtaatgaaaggatataacccatlvggcagcgatlvgtatataaagaccgcvgaaactatgcattc | 123 |
| Db | 154 | gtcatgaaaggtattataaccatgycacatgcttvgtatataagacaacgcttataacagctatt | 213 |
| QY | 124 | gaagaggattgcaaatataccggtgcgtcaataaaggtccggagtgtgtatctatgvtggggacaa | 183 |
| Db | 214 | ccttcacattgcaagacgaaggcgcacacacacgattcttgatctgtcttatacagaattgcvgccaa | 273 |
| QY | 184 | lvgaacaagaatlvgaatcccatcacatgataagaacattatctcttttaacggaagataatcat | 243 |
| Db | 274 | lvggaaanaaagacgcacattgacacacattcgtgaattcatattgactgtgvggcgaataaana | 333 |
| QY | 244 | ttggtgtcgtcttcttgaagtccatgatagctacgcggttatatgattccattgcttcgcctcat | 303 |
| Db | 334 | atvgctvgcgttcgvtcttgaagtccatgatagctacgcggttcgcgtattcgcgaatgatttaaa | 393 |
| QY | 304 | cgctcgtctgatactatgtgatlvgaatbagaaggtctcttaattggaagaagagataacgctc | 363 |
| Db | 394 | cgaagcgttggtattatgtgatbagaatgaaagatbvgcgttatcgtttaagaagaagatacgcgtc | 453 |
| QY | 364 | attattaatatcgcgaatagatggttttggctcgtlvggaaagvggaatgcttvggcgtacagg | 423 |
| Db | 454 | attatcaecatlvgaacacgaatggtgtacatvggaaagtvtgvgatvgctacagcttvgccgaatgvc | 513 |
| QY | 424 | tataacaagaacatcccgcgagctvgtaacgcgcgttaaacacatcattgattgattagat | 483 |
| Db | 514 | tatatgtatgcatctccgaagcttcgcgagctgcgcggttaacacacacacttaacgtttgat | 573 |
| QY | 484 | gctcvcgggggtvgggacaattccacaatcgaattcatatgatatvgaaagaaagaattttaat | 543 |
| Db | 574 | gcacacagagatbvgggcacaataaccgaactcatatcatgatatggaacagaagtgcgtttaat | 633 |
| QY | 544 | gctacacctcaagaataatacatgatttttgcattcatatgatatgtaataatgaaagtgtgtaat | 603 |
| Db | 634 | gcgaatccggttaaaaaataacgaatgcttccctccatccatgatactgatagtcgttvggtgat | 693 |
| QY | 604 | gcacatcgaagtctgtaataatattgacacgcggtttcttaatacaagacctgcgattatcaatt | 663 |
| Db | 694 | gctaacactgttagatacaaaatattgtataggtctatagatacaagacctgttcctcgttaata | 753 |
| QY | 664 | ggtgaaatttggaccgcgtcatatacaaaatggtgaaagtgcgaatgaagcaacgattatgagctat | 723 |

| | | | | | |
|-------------------------------------------|---|---|-----|---------------------------------------------------------------------------|---------------------------------------------------|
| D | b | | 754 | ggtgaattcggctcatagacatactgatgtgtagtgatgagaatcatcaacttcttgtttt | 813 |
| O | y | | 724 | tctcaacaagaaggagtctgggtgtgttcggtgcgtgtgtlcatlgaaaggagcccggaatyg | 783 |
| D | b | | 814 | tctgaagaacctggcacagaggtgtgtctgcgttgcgttcttgtaagycacacagttaccgatatg | 873 |
| O | y | | 784 | gagattttaaaccttcgtaatgttatgttgctgcggaaaataactccaacttcagcttgggggaatta | 843 |
| D | b | | 874 | gacctattgaacctctttagaagaactgggtcgtgcgcacacatttaactgtattcgggggaataga | 933 |
| O | y | | 844 | atagtgaatggtcatatggtttaagaagaaccttcagattcaagccacglttttaca | 900 |
| D | b | | 934 | attgccacggggcggatggtcttacagagaacctccaacacatccacgcgtatttaca | 990 |
| <hr/> | | | | | |
| RESULT 6 | | | | | |
| D | R | I | D | Z45337 standard; DNA; 1482 BP. | |
| X | X | A | C | Z45337; | |
| X | X | D | E | 27-MAR-2000 (first entry) | |
| DNA encoding a Bacillus mannanase enzyme. | | | | | |
| X | X | M | N | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | |
| X | X | K | M | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | |
| X | X | K | M | mannan; galactomannan; glucomannan; galactoglucomannan; celluloseic fibre | |
| X | X | K | M | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | |
| X | X | K | M | plant material degradation; recycled waste paper; paper making pulp; | |
| X | X | K | M | glue; locust bean gum; mannan-containing food; coffee extract; | |
| X | X | K | M | cleaning composition; machine washing; hard-surface cleaner; | |
| X | X | K | M | dishwashing; oral; dental; contact lens; body-care composition; | |
| X | X | S | O | fabric softener; oil well drilling; subterranean formation fracture; ss. | |
| <hr/> | | | | | |
| X | X | B | a | Bacillus agaradhaerens. | |
| F | H | K | E | Key | Location/Qualifiers |
| F | T | C | D | CDS | 1..1482 |
| F | T | / | a | /*tag- | |
| F | T | / | - | product= | "mannanase" |
| F | T | 1 | . | 93 | |
| F | T | / | * | tag- | b |
| <hr/> | | | | | |
| X | N | | | W09964619-A2. | |
| X | P | | | 16-DEC-1999. | |
| X | P | | | 10-JUN-1999; | '99WO-DK00314. |
| X | X | | | 10-JUN-1998; | '98US-0111256. |
| X | P | | | 20-OCT-1998; | '98DK-0001340. |
| X | P | | | 20-OCT-1998; | '98DK-0001341. |
| X | P | | | 28-OCT-1998; | '98US-0105970. |
| X | P | | | 28-OCT-1998; | '98US-0106054. |
| X | P | | | 23-DEC-1998; | '98DK-0001725. |
| X | P | | | 05-MAR-1999; | '99DK-0000306. |
| X | P | | | 05-MAR-1999; | '99DK-0000307. |
| X | P | | | 05-MAR-1999; | '99DK-0000308. |
| X | P | | | 05-MAR-1999; | '99DK-0000309. |
| X | P | | | 09-MAR-1999; | '99US-0123543. |
| X | P | | | 10-MAR-1999; | '99US-0123623. |
| X | P | | | 10-MAR-1999; | '99US-0123641. |
| X | P | | | 11-MAR-1999; | '99US-0123642. |
| <hr/> | | | | | |
| (NOVO) NOVO-NORDISK AS. | | | | | |
| X | P | | | Kauppinen MS, | Schuelein M, Schnorr K, Andersen LN, Bjornvad ME; |
| X | X | | | WPI: 2000-105891/09. | |
| D | R | | | P-PSDB; Y54124. | |
| X | X | | | | |

PI Kauppinen MS, Schneelein M, Schnorr K, Andersen LN, Bjornvad ME, X
XX
DR WPI: 2000-105891/09.
DR P-PSDB: Y54124.
XX

PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions

Example 5; Page 212-213; 242pp; English.

CC The present sequence encodes a *Bacillus mannanase* (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
 CC mannanase hydrolyses galactomannans. Specifically, mannases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
 CC and galactoglucomannans. The mannanase protein, or preparations
 CC containing it, are used to improve properties of cellulosic or
 CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.

CC Sequence 1482 BP; 446 A; 285 C; 352 G; 399 T; 0 other;

Query Match 52.0%; Score 468.2; DB 21; Length 1482;

Best Local Similarity 70.1%; Pred. No. 2,1e-131;

Matches 629; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

OY 4 gcaattccggaattattatgtaagcggtaccatctatcatcgatgcaatggaaccattt 63

DB 94 gcaattccggaattattatgtaagcggtaccatctatcatcgatgcaatggaaccattt 153

OY 64 gtaattccggaattattatgtaagcggtaccatctatcatcgatgcaatggaaccattt 123

DB 154 gtaattccggaattattatgtaagcggtaccatctatcatcgatgcaatggaaccattt 213

OY 124 gaaaggaattgcaataaccggtgctgtaacggtcgcggtatgctgtaacggtggaacaa 183

DB 214 cctgcatctgcaagagcaagcgcccaacggtatgctgtaacggtggaacaa 273

OY 184 tggacaagaatgacatccatcagtaagaaccttctcttgcggaagataatcat 243

DB 274 tggacaagaatgacatccatcagtaagaaccttctcttgcggaagataatcat 333

OY 244 ttggtgctgcttgaagatgacatcagtcggttgcggtatgctgtaacggtggaacaa 303

DB 334 atggtgctgcttgaagatgacatcagtcggttgcggtatgctgtaacggtggaacaa 393

OY 304 cgtgctgcttgaagatgacatcagtcggttgcggtatgctgtaacggtggaacaa 363

DB 394 cgaagcgttgaatgataagaatgaaagtgctgtaacggtggaacaa 453

OY 364 attattaattcgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 423

DB 454 attattaattcgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 513

OY 424 tataaacaagaatccggaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 483

DB 514 tataaacaagaatccggaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 573

OY 484 gctgcggtggtggaatccggaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 543

DB 574 gcaagcgttgaatgataagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 633

OY 544 gctgaacctcaagaatacaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 603

DB 634 gcaagcgttgaatgataagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 693

OY 604 gcaatcgaatgctgtaacatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 663

DB 694 gctaacactgtatgatacaaatgataagtagatgataagtagatgataagtagatgataagtagatgata 753

OY 664 gctgaattggaacacgctgatacaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 723

DB 754 gctgaattggaacacgctgatacaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 813

OY 724 tctgaacaagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 783

DB 814 tctgaacaagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 873

OY 784 gactattagacacacgctgatacaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 843

DB 874 gactattagacacacgctgatacaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 933

OY 844 atagtaagatgcaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 900

DB 934 atgtaagatgcaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 990

RESULT 7

ID 245342 standard; DNA; 1107 BP.

AC 245342;

DT 27-MAR-2000 (first entry)

DE DNA encoding a *Bacillus* sp. AA349 mannanase enzyme.

XX

XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;

XX endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;

XX mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;

XX synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;

XX plant material degradation; recycled waste paper; paper making pulp;

XX guar; locust bean gum; mannan-containing food; coffee extract;

XX cleaning composition; machine washing; hard-surface cleaner;

XX dishwashing; oral; dental; contact lens; body-care composition;

XX fabric softener; oil well drilling; subterranean formation fracture; ss.

XX *Bacillus* sp.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;

XX (NOVO) NOVO-NORDISK AS.
 PA Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 PI MPI: 2000-105891/09.
 XX P-PSDB: Y54127.
 DR
 XX New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PS Example 10; Page 219-220; 242pp; English.
 XX
 CC The present sequence encodes a *Bacillus mannanase* (also known as
 CC Mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 SQ Sequence 995 BP; 286 A; 162 C; 256 G; 291 T; 0 other;

Query Match 45.0%; Score 405; DB 21; Length 995;
 Best Local Similarity 66.1%; Pred. No. 1,9e-112;
 Matches 585; Conservative 0; Mismatches 300; Indels 0; Gaps 0;

QY 13 ggaattatgtaagcgggtaccactctatcagatgcagatggaacccattgtatgaga 72
 DB 106 gggcttcattatgatacagcattatgatacgaatggaaccccttgtatgaga 165
 QY 73 gggatatacattgagcgcgcatgataaagcagcgaactctgtcaattgagagatt 132
 DB 166 gggatatacattgagcgcgcatgataaagcagcgaactctgtcaattgagagatt 225
 QY 133 gcaataccggtctaatagatgctcggatgtatctgatgggagacatgacaaa 192
 DB 226 agtcaaacaggggcaataatagatcgtctgttcttgcataagggcaaatgacaaa 285
 QY 193 gatgacatcacaacgaagaacactatctcttagcggagaataatctgtgtct 252
 DB 286 gatgacatcacaacgaagaacactatctcttagcggagaataatctgtgtct 345
 QY 253 gtctctgaattatgctgtacccgtgtatgattccattcttcgcgaatcgtgtct 312
 DB 346 gtctctgaattatgctgtacccgtgtatgattccattcttcgcgaatcgtgtct 405
 QY 313 gattatggaatgaaagagagatgcttatttgaagagggaatacgtcatattaat 372
 DB 406 gactattggaatgaaagagagatgcttatttgaagagggaatacgtcatattaat 465
 QY 373 attgcgaatgaatggttgcgttcgggaaaggagatgcttgagcgggtataaaca 432
 DB 466 atgcgaatgaatggttgcgttcgggaaaggagatgcttgagcgggtatacaga 525
 QY 433 gcaatcccgatgctgtaacgcgcggtataacacattcttgatggatgagatgcgggg 492
 DB 526 ggcatacgtcagcttcgaatgcaggcttgcacatacatctttagcgcgtgcgggt 585
 QY 493 tgggacaatttccacacatgattcatgattatggaagagagattttagtgcgacct 552

DB 586 tatggccagtaaccctcaatcgtgattgattatgtaaggaattataatgctgacca 645
 QY 553 caacgaatacacaatgcttcttcattcatatgataatgataatgataatgataatgata 612
 DB 646 cagaagaacacaaatgcttcttcattcatatgataatgataatgataatgataatgata 705
 QY 613 gtctctgaattatgctgtacccgtgtatgattccattcttcgcgaatcgtgtct 672
 DB 706 gtaagacgaacacattgctgtacccgtgtatgattccattcttcgcgaatcgtgtct 765
 QY 673 ggaacccgctacacaaatgctgtacccgtgtatgattccattcttcgcgaatcgtgtct 732
 DB 766 gggcattgcatatgatacgggtgtatgataaggaacacattttagctattccagcaa 825
 QY 733 agagagattggtggttgcgtgtgctgatacgaagggagacggccagatgattta 792
 DB 826 agaatgtggaattggttgcgtgtgctgatacgaagggagacggccagatgattta 885
 QY 793 gactctggaatgattggttgcgtgtgctgatacgaagggagacggccagatgattta 852
 DB 886 gattatcgaatgacttgcgtgtgctgatacgaagggagacggccagatgattta 945
 QY 853 gtcacatgatttaagagaacttcgagattagacacgctttt 897
 DB 946 gttccgaatggtgatttgcataacacctcaagaagacggtgttt 990

RESULT 9
 245345
 ID 245345 standard; DNA; 960 BP.
 XX
 AC 245345;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE DNA encoding a *Bacillus* sp. mannanase enzyme.

OS *Bacillus* sp.
 OS
 FH Key Location/Qualifiers
 FT CDS 1..960
 FT /tag- a
 FT /product- "mannanase"
 FT /note- "not termination codon given"
 FT sig_peptide 1..84
 FT /tag- b

PN MO9964619-A2.
 PN
 PD 16-DEC-1999.
 PD
 PE 10-JUN-1999; 99MO-DK00314.
 PE
 XX 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.

PR 05-MAR-1999: 99DK-0000308.
 PR 05-MAR-1999: 99DK-0000309.
 PR 09-MAR-1999: 99US-0123543.
 PR 10-MAR-1999: 99US-0123623.
 PR 10-MAR-1999: 99US-0123641.
 PR 11-MAR-1999: 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 PI Kauplinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 DR P-PSDB: Y54133.
 XX
 PT New mannanses for treatment of textiles, plant material and coffee
 XX extract, and in cleaning compositions
 PS
 XX Disclosure; Page 229; 242pp; English.

XX The present sequence encodes a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.

XX Sequence 564 BP; 159 A; 111 C; 158 G; 136 T; 0 other;

Query Match 18.6%; Score 167.8; DB 21; Length 564;
 Best Local Similarity 59.2%; Pred. No. 4.8e-41;
 Matches 286; Conservative 0; Mismatches 197; Indels 0; Gaps 0;
 QY 3 tgcgaatttcggaatttgaagtcggtacacactatacagatgcgaatcccat 62
 DB 81 tgcacaagaagcgcttcacgtaaaagtcagagttgttgacaaaatgsgcgtctta 140
 QY 63 tgcgaatgagaggtatcaccatgagcgcagatgataaagaccaggaactactgcaat 122
 DB 141 cgttatgctggtggtcgcacacatgacatctctgtttacaagaatttaagagagcaat 200
 QY 123 tgaagagattgcgaatccggtcttaacagtcgcgcagattgtgtatctatgagggaca 182
 DB 201 cccctgcacatgcgaagaacagggcgacacacagtgagatagctcttaacatgacagca 260
 QY 183 atggacaagaatgacatcacaacagtaagaacactatctctttagcggagaagataca 242
 DB 261 atggagaagaatgacatccctcgtgctgcgcgtgtgctgtccacagaaacatatg 320
 QY 243 ttggtgtcgtctcgtgaatgtatgctacgcgtgtatgattcattccttcgtccaa 302
 DB 321 gtgcacaacgctgtgaaatgcacagatgtacaggaagtgaatcccatgattaga 380
 QY 303 tctgtctgttattgattggaatgagaagtgcttattattggaaggaagatacgt 362
 DB 381 taagacagctgattactggtcgaatgagtgctgtatgtcttaaggggacaaagccggt 440
 QY 363 catattaatatgcgaatgaatggtttgttctgttggaagggaggtcttggtgcagc 422
 DB 441 aatcattaatatgcgaatgaatggtttgttctgttggaagggaggtcttggtgcagc 500

QY 423 gtataacaagaacatcccgatgtgcgttaacgctcttaacacattatgtag 482
 DB 501 atacgcacaagcgtatcccgcttgcgagtgctgtgcctgcacatacgttaataatga 560
 QY 483 tgc 485
 DB 561 tgc 563

RESULT 11

245344
 245344 standard; DNA; 397 BP.

XX 245344;

DT 27-MAR-2000 (first entry)

XX DNA encoding a partial *Bacillus* mannanase enzyme.

XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture; ss.

XX *Bacillus* clausii.

PH Key Location/Qualifiers

FT CDS 1..396

FT /tag- a

FT /product- "mannanase"

FT /note- "partial sequence"

XX W09964619-A2.

XX 16-DEC-1999.

XX 10-JUN-1999; 99WO-DK00314.

XX 10-JUN-1998; 98US-0111256.

XX 20-OCT-1998; 98DK-0001340.

XX 20-OCT-1998; 98DK-0001341.

XX 28-OCT-1998; 98US-0105970.

XX 28-OCT-1998; 98US-0106054.

XX 23-DEC-1998; 98DK-0001725.

XX 05-MAR-1999; 99DK-0000306.

XX 05-MAR-1999; 99DK-0000307.

XX 05-MAR-1999; 99DK-0000308.

XX 05-MAR-1999; 99DK-0000309.

XX 09-MAR-1999; 99US-0123543.

XX 10-MAR-1999; 99US-0123623.

XX 10-MAR-1999; 99US-0123641.

XX 11-MAR-1999; 99US-0123642.

XX (NOVO) NOVO-NORDISK AS.

PI Kauplinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;

XX WPI: 2000-105891/09.

XX P-PSDB: Y54131.

XX

PT New mannanses for treatment of textiles, plant material and coffee

XX extract, and in cleaning compositions

PS Disclosure; Page 226-227; 242pp; English.

XX

XX The present sequence encodes a mannanase enzyme (also known as

CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).

XX The mannase hydrolyses galactomannans. Specifically, mannases

| | | |
|----|---------------|--------------------------|
| FT | | /product-"mannanase" |
| FT | | /note="partial sequence" |
| XX | | |
| XX | WO9964619-A2. | |
| XX | | |
| PD | 16-DEC-1999. | |
| XX | | |
| PF | 10-JUN-1999; | 99WO-DK00314. |
| XX | | |
| PR | 10-JUN-1998; | 98US-0111256. |
| PR | 20-OCT-1998; | 98DK-0001340. |
| PR | 20-OCT-1998; | 98DK-0001341. |
| PR | 28-OCT-1998; | 98US-0105870. |
| PR | 28-OCT-1998; | 98US-0106054. |
| PR | 23-DEC-1998; | 98DK-0001725. |
| PR | 05-MAR-1999; | 99DK-0000306. |
| PR | 05-MAR-1999; | 99DK-0000307. |

PR 03-MAR-1999; 39DA-000509.
PR 09-MAR-1999; 99US-012543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 10-MAR-1999; 99US-0123641.

PK 11-MAR-1999; 9905-0123642.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX

PI Kaupplinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
DR WPI; 2000-105891/09.
DB P-DEMB. VS4130

XX New mannanases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
PT

PS Disclosure; page 225; 242pp; English.
XX
CC The present sequence encodes a mannase enzyme (also known as

CC The mannanase hydrolyses galactomannans. Specifically, mannanases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans
CC glucomannans, and galactoglucomannans. The mannanase protein, or

plant materials (particularly recycled waste paper, paper making pulps or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-sides or printing pastes). They are also used to degrade or modify

CC are also used to process coffee extracts (to inhibit gel formation)
CC cleaning compositions (for machine washing of fabrics, as hard-surf
CC reduce viscosity of mannan-containing foods or feeds). The mannans
CC or material containing guar or locust bean gums (chickeners), or to
CC

CC softeners. They can also be used in all well drilling to fracture
CC soils and prevent binding of some soils to cellulose; and in fabr
CC lens or body-care compositions) where they remove mannan-containing
CC cleaners, for hand or machine disinfecting, also in oral, dental, co

| | |
|----|-------------------------------------------------------|
| CC | subterranean formations, |
| XX | |
| SQ | Sequence 915 BP; 267 A; 203 C; 229 G; 216 T; 0 other; |

| | | | | |
|-----------------------|--------|--------------------|--------|-------------|
| Query Match | 8.7%; | Score 78.2; | DB 21; | Length 915; |
| Best Local Similarity | 52.8%; | Pred. No. 6.1e-14; | | |
| Matched | 354 | Conserved | 0 | Matched |

436 atccgcgattgcgtaacgcgcgctcaaacatattgtagatgctgcggggtgg 4

QY 496 ggacaattccacaatcgattcatgattatggagagaagtlttaatgctgcacctaa 5
||||| | ||||| | ||||| | ||||| | ||||| |

Qy 556 cgaataacaatgttttcgattcatatgtatcgataatgcaggtcgttaatgatcgcgaagtt 6

[illegible]

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 09:56:10 ; Search time 75.74 Seconds
(without alignments)
1797.160 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990
Perfect score: 900
Sequence: 1 aatgcaaatcgcgatttta.....gattgaagcaccgttttaca 900

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCrus.COMB.seq:*
7: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| c 1 | 37.4 | 4.2 | 7218 | 1 US-08-232-463-14 | Sequence 14, Appl |
| c 2 | 33.4 | 3.7 | 1365 | 4 US-08-870-827-4 | Sequence 4, Appl |
| c 3 | 33.4 | 3.7 | 2408 | 4 US-08-870-827-5 | Sequence 5, Appl |
| c 4 | 33.4 | 3.7 | 1800 | 5 US-09-039-773A-1 | Sequence 1, Appl |
| c 5 | 32 | 3.6 | 4359 | 3 US-08-566-398-23 | Sequence 23, Appl |
| c 6 | 32 | 3.6 | 4365 | 6 PCT-US91-08525-21 | Sequence 21, Appl |
| c 7 | 32 | 3.6 | 4365 | 6 PCT-US91-08525-25 | Sequence 25, Appl |
| c 8 | 32 | 3.6 | 4365 | 6 PCT-US93-04384-1 | Sequence 1, Appl |
| c 9 | 32 | 3.6 | 4365 | 6 PCT-US93-04384-7 | Sequence 7, Appl |
| c 10 | 32 | 3.6 | 4365 | 6 PCT-US93-04384-15 | Sequence 15, Appl |
| c 11 | 32 | 3.6 | 6090 | 3 US-08-566-398-35 | Sequence 35, Appl |
| c 12 | 32 | 3.6 | 6144 | 3 US-08-566-398-32 | Sequence 32, Appl |
| c 13 | 31.8 | 3.5 | 7366 | 7 5169760-3 | Patent No. 5169760 |
| c 14 | 31.8 | 3.5 | 7852 | 5 US-08-836-022A-2 | Sequence 2, Appl |
| c 15 | 31.8 | 3.5 | 7897 | 5 US-08-836-022A-1 | Sequence 1, Appl |
| c 16 | 31.8 | 3.5 | 8509 | 2 US-08-462-014-1 | Sequence 1, Appl |
| c 17 | 31.8 | 3.5 | 10398 | 3 US-08-331-384-1 | Sequence 1, Appl |
| c 18 | 31.8 | 3.5 | 10398 | 3 US-08-708-188-1 | Sequence 1, Appl |
| c 19 | 31.8 | 3.5 | 10398 | 3 US-08-836-087-1 | Sequence 1, Appl |
| c 20 | 31.8 | 3.5 | 19182 | 4 US-08-850-880-11 | Sequence 11, Appl |
| c 21 | 31.8 | 3.5 | 19182 | 4 US-08-944-916-11 | Sequence 11, Appl |
| c 22 | 31 | 3.4 | 1453 | 1 US-08-252-995D-1 | Sequence 1, Appl |
| c 23 | 31 | 3.4 | 1453 | 1 US-08-834-108-1 | Sequence 1, Appl |
| c 24 | 31 | 3.4 | 1600 | 4 US-08-252-995D-5 | Sequence 5, Appl |
| c 25 | 31 | 3.4 | 1600 | 4 US-08-834-108-5 | Sequence 5, Appl |
| c 26 | 31 | 3.4 | 2882 | 1 US-08-424-567-1 | Sequence 1, Appl |

| | | | | | |
|------|------|-----|-------|--------------------|-------------------|
| 27 | 31 | 3.4 | 2882 | 3 US-08-711-928-1 | Sequence 1, Appl |
| 28 | 31 | 3.4 | 3447 | 1 US-08-252-995D-3 | Sequence 3, Appl |
| 29 | 31 | 3.4 | 3447 | 4 US-08-834-108-3 | Sequence 3, Appl |
| c 30 | 30.6 | 3.4 | 18994 | 2 US-08-459-586-4 | Sequence 4, Appl |
| c 31 | 30.6 | 3.4 | 18994 | 2 US-08-282-696-4 | Sequence 4, Appl |
| c 32 | 30.2 | 3.4 | 5718 | 5 US-08-714-918-48 | Sequence 48, Appl |
| c 33 | 30 | 3.3 | 2675 | 1 US-07-749-001-4 | Sequence 4, Appl |
| c 34 | 30 | 3.3 | 2675 | 1 US-08-154-188-4 | Sequence 4, Appl |
| c 35 | 30 | 3.3 | 2675 | 1 US-08-463-335-4 | Sequence 4, Appl |
| c 36 | 30 | 3.3 | 2675 | 3 US-08-464-023A-4 | Sequence 4, Appl |
| c 37 | 30 | 3.3 | 5319 | 2 US-08-169-927-1 | Sequence 1, Appl |
| c 38 | 29.8 | 3.3 | 1029 | 5 US-08-978-456-1 | Sequence 1, Appl |
| c 39 | 29.8 | 3.3 | 3336 | 3 US-08-978-456-7 | Sequence 7, Appl |
| c 40 | 29.8 | 3.3 | 3336 | 3 US-08-978-456-7 | Sequence 7, Appl |
| c 41 | 29.8 | 3.3 | 4467 | 2 US-08-565-907A-1 | Sequence 1, Appl |
| c 42 | 29.8 | 3.3 | 4467 | 4 US-08-910-551B-1 | Sequence 1, Appl |
| c 43 | 29.8 | 3.3 | 4467 | 4 US-08-909-425A-1 | Sequence 1, Appl |
| c 44 | 29.8 | 3.3 | 51952 | 5 US-08-947-823-1 | Sequence 1, Appl |
| c 45 | 29.6 | 3.3 | 1716 | 5 US-09-029-267-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTgpt-Fls
US-08-232-463-14

Query Match 4.2%; Score 37.4; DB 1; Length 7218;
Best Local Similarity 5.7%; Pred. No. 0.068;
Matches 20; Conservative 180; Mismatches 151; Indels 0; Gaps 0;

[illegible]

RESULT 2
US-08-870-827-4

```

Sequence 4, Application US/08870827
Patent No. 5962297
GENERAL INFORMATION:
APPLICANT: Tsusaki et al
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOFURANOSIDASE
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEO ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1365 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
IS-08-870-827-4

```

| | | | | |
|-----------------------|-----------------|-----------------|-----------|--------------|
| Query Match | 3.7% | Score 33.4; | DB 4; | Length 1365; |
| Best Local Similarity | 55.7%; | Pred. NO. 0.53; | | |
| Matches 64; | Conservative 0; | Mismatches 51; | Indels 0; | Gaps 0; |

OY 172 gaaggagacaaatggacaagaatgacatcccatcagtaagaaccttactcttgcg 23
Dd 655 gatgaagccaagtggattttcagcgcataaacatctttaaagaccctcactactgtgaa 71
OY 232 gaagataatcattgtgtcgtcgtcttcgaagtcaagatgcacggttatgatt 286
Db 715 gataagggccctaataatatcttgttttgaagccgaatctggaamaacagagatgtt 769

RESULT 3
US-08-870-827-F

```

Sequence 5, Application US/08870827
Patent No. 5962297
GENERAL INFORMATION:
APPLICANT: Tsusaki et al.
TITLE OF INVENTION: POLYPEPTIDE HAVING -FRUCTOSE
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROMDY AND NEWMARK
STREET: 419, Seventh Street, N.W., Suite 300
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,827
FILING DATE: 06-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 170,630/1996
FILING DATE: 10-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: BROMDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEO ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2408 base pairs
type: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Bacillus sp.
INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..360
IDENTIFICATION METHOD: E
NAME/KEY: signal peptide
LOCATION: 361..456
IDENTIFICATION METHOD: S
NAME/KEY: mat peptide
LOCATION: 457..1821
IDENTIFICATION METHOD: S
NAME/KEY: 3'UTR
LOCATION: 1822..2408
IDENTIFICATION METHOD: E

```



```

RESULT      6
PCT-US91-08525-21
; Sequence 21, Application PC/TUS9108525
GENERAL INFORMATION:
APPLICANT: SmithKline Beecham, Corporation
TITLE OF INVENTION: Recombinant feline Coronavirus S
TITLE OF INVENTION: Proteins
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2739
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08525
FILING DATE: 19911114
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Kling, William T.
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: SBC 14532B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO.: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US91-08525-21

Query Match          3.6%; Score 32; DB 6; Length 4365;
Best local similarity 65.3%; Pred. No. 2.6; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 25;

QY 520 gattatggaaagaaagtttttaaqtctgacccccaacgaatacatgtttcgattcat 579
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2593 gatgtgcgaagaactgttgtaatacgttaaccctaagatgtaaacaaattgtttaacacaatat 2652
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 580 atgtatgtaaat 591
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2653 gtgtctgcattg 2664

RESULT      7
PCT-US91-08525-25
; Sequence 25, Application PC/TUS9108525
GENERAL INFORMATION:
APPLICANT: SmithKline Beecham, Corporation
TITLE OF INVENTION: Recombinant feline Coronavirus S
TITLE OF INVENTION: Proteins
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
```

```

1 STREET: 709 Swedeland Road
2 CITY: King of Prussia
3 STATE: PA
4 COUNTRY: USA
5 ZIP: 19406-2799
6
7 COMPUTER READABLE FORM:
8
9 MEDIUM TYPE: Floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: Patent Release #1.0, Version #1.25
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: PCT/US91/08525
16 FILING DATE: 19911114
17 CLASSIFICATION: 435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 07/698,927
20 FILING DATE: 13-MAY-1991
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 07/613,066
23 FILING DATE: 14-NOV-1990
24 ATTORNEY/AGENT INFORMATION:
25 NAME: King, William T.
26 REGISTRATION NUMBER: 30,954
27 REFERENCE/DOCKET NUMBER: SPC 14532B
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (215) 270-5015
30 TELEFAX: (215) 270-5090
31 INFORMATION FOR SEQ ID NO: 25:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 4365 base pairs
34 TYPE: NUCLEIC ACID
35 STRANDEDNESS: double
36 TOPOLOGY: unknown
37 MOLECULE TYPE: CDNA
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: 1..4362
41 PCT-US91-08525-25
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384
FILING DATE: 19930507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SBC H85009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US93-04384-1

```

```

Query Match          3.68; Score 32; DB 6; Length 4365;
Best Local Similarity 65.38; Pred. No. 2.6;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

```

QY 520 gattatggaagaagttttaatgctgcaccccaacgaataatattgattcat 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2593 GATTGTGCAGAGATACGTTTGTAATGCTACCCCTAGATGTAAACAATTTGTTAACACAAATAT 2652
QY 580 atgtatgaatat 591
    ||||| |||||
DB 2653 GTGCTGCATCT 2664

```

```

RESULT 9
PCT-US93-04384-7
Sequence 7, Application PC/TUS9304384
GENERAL INFORMATION:
APPLICANT: Miller, Timothy J.
APPLICANT: Klepfer, Sharon
APPLICANT: Reed, Albert Paul
APPLICANT: Jones, Elaine V.
TITLE OF INVENTION: Compositions and Methods for Vaccination
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation - Corporate
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384
FILING DATE: 19930507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SBC H85009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US93-04384-7

```

```

Query Match          3.68; Score 32; DB 6; Length 4365;
Best Local Similarity 65.38; Pred. No. 2.6;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```

```

QY 520 gattatggaagaagttttaatgctgcaccccaacgaataatattgattcat 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2593 GATTGTGCAGAGATACGTTTGTAATGCTACCCCTAGATGTAAACAATTTGTTAACACAAATAT 2652
QY 580 atgtatgaatat 591
    ||||| |||||
DB 2653 GTGCTGCATCT 2664

```

```

RESULT 10
PCT-US93-04384-15
Sequence 15, Application PC/TUS9304384
GENERAL INFORMATION:
APPLICANT: Miller, Timothy J.
APPLICANT: Klepfer, Sharon
APPLICANT: Reed, Albert Paul
APPLICANT: Jones, Elaine V.
TITLE OF INVENTION: Compositions and Methods for Vaccination
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation - Corporate
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384

```

FILING DATE: 19930507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SBC H85009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 4365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4362
PCT-US93-04384-15

Query Match
Best Local Similarity 65.3%; Score 32; DB 6; Length 4365;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 520 gattatggaagaagtttttaatgctgacccccaacgaataatgtttcgattcat 579
|||||
Db 2593 GATTGTGCAAGATACGTTGTAAAGTAACCTAGATGTACAAATGTTAACCAATAT 2652

QY 580 atgatatgaatat 591
|||||
Db 2653 GTGTCTGCATGT 2664

RESULT 11
US-08-566-398-35
Sequence 35, Application US/08566398
Patent No. 5858373
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIOUS
PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
TITLE OF INVENTION: MAKING AND USING THEM
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/566,398
FILING DATE: 01-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2880
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 6090 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-566-398-35

Query Match
Best Local Similarity 65.3%; Score 32; DB 3; Length 6090;
Matches 47; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 520 gattatggaagaagtttttaatgctgacccccaacgaataatgtttcgattcat 579
|||||
Db 3123 GATTGTGCAAGATACGTTGTAAAGTAACCTAGATGTACAAATGTTAACCAATAT 3182

QY 580 atgatatgaatat 591
|||||
Db 3183 GTGTCTGCATGT 3194

RESULT 12
US-08-566-398-32
Sequence 32, Application US/08566398
Patent No. 5858373
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIOUS
PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
TITLE OF INVENTION: MAKING AND USING THEM
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/566,398
FILING DATE: 01-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2880
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 6144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-566-398-32

Query Match
3.6%; Score 32; DB 3; Length 6144;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2000, 01:58:22 ; Search time 1967.85 Seconds
(without alignments)
2827.718 Million cell updates/sec

Title: US-09-339-159-1_COPY_91_990
Perfect score: 900
Sequence: 1 aatgcaaatccgatttta.....gattaagcacgctttta 900

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 segs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

| | |
|-----|--------------|
| 1: | gb_est1:* |
| 2: | gb_est2:* |
| 3: | gb_est3:* |
| 4: | gb_est4:* |
| 5: | gb_est5:* |
| 6: | gb_est6:* |
| 7: | gb_est7:* |
| 8: | gb_est8:* |
| 9: | gb_est9:* |
| 10: | gb_est10:* |
| 11: | gb_est11:* |
| 12: | gb_est12:* |
| 13: | gb_est13:* |
| 14: | gb_est14:* |
| 15: | gb_est15:* |
| 16: | gb_est16:* |
| 17: | gb_est17:* |
| 18: | gb_est18:* |
| 19: | gb_est19:* |
| 20: | gb_est20:* |
| 21: | gb_est21:* |
| 22: | gb_est22:* |
| 23: | gb_est23:* |
| 24: | gb_est24:* |
| 25: | gb_est25:* |
| 26: | gb_est26:* |
| 27: | gb_est27:* |
| 28: | gb_est28:* |
| 29: | gb_est29:* |
| 30: | gb_est30:* |
| 31: | gb_est31:* |
| 32: | gb_est32:* |
| 33: | gb_est33:* |
| 34: | gb_est34:* |
| 35: | gb_est35:* |
| 36: | gb_est36:* |
| 37: | gb_est37:* |
| 38: | gb_est38:* |
| 39: | gb_est39:* |
| 40: | gb_est40:* |
| 41: | em_est4a:* |
| 42: | em_estfun:* |
| 43: | em_esthum1:* |

| | |
|------|---------------|
| 44: | em_esthum2:* |
| 45: | em_esthum3:* |
| 46: | em_esthum4:* |
| 47: | em_esthum5:* |
| 48: | em_esthum6:* |
| 49: | em_esthum7:* |
| 50: | em_esthum8:* |
| 51: | em_esthum9:* |
| 52: | em_esthum10:* |
| 53: | em_esthum11:* |
| 54: | em_esthum12:* |
| 55: | em_esthum13:* |
| 56: | em_esthum14:* |
| 57: | em_esthum15:* |
| 58: | em_esthum16:* |
| 59: | em_esthum17:* |
| 60: | em_esthum18:* |
| 61: | em_esthum19:* |
| 62: | em_esthum20:* |
| 63: | em_estlin1:* |
| 64: | em_estlin2:* |
| 65: | em_estlin3:* |
| 66: | em_estlin4:* |
| 67: | em_estom:* |
| 68: | em_estov1:* |
| 69: | em_estov2:* |
| 70: | em_estp11:* |
| 71: | em_estp12:* |
| 72: | em_estp13:* |
| 73: | em_estp14:* |
| 74: | em_estp15:* |
| 75: | em_estro1:* |
| 76: | em_estro2:* |
| 77: | em_estro3:* |
| 78: | em_estro4:* |
| 79: | em_estro5:* |
| 80: | em_estro6:* |
| 81: | em_estro7:* |
| 82: | em_estro8:* |
| 83: | em_estro9:* |
| 84: | em_estro10:* |
| 85: | em_estro11:* |
| 86: | em_estro12:* |
| 87: | em_estro13:* |
| 88: | gb_gss1:* |
| 89: | gb_gss2:* |
| 90: | gb_gss3:* |
| 91: | gb_gss4:* |
| 92: | em_gss1:* |
| 93: | em_gss2:* |
| 94: | em_gss3:* |
| 95: | em_gss4:* |
| 96: | gb_gss5:* |
| 97: | gb_gss6:* |
| 98: | gb_gss7:* |
| 99: | gb_gss8:* |
| 100: | gb_gss9:* |
| 101: | em_gss5:* |
| 102: | em_gss6:* |
| 103: | em_gss7:* |
| 104: | em_gss8:* |
| 105: | em_gss9:* |
| 106: | em_gss10:* |
| 107: | em_gss11:* |
| 108: | gb_gss10:* |
| 109: | gb_gss11:* |
| 110: | em_gss12:* |
| 111: | gb_gss12:* |
| 112: | gb_gss13:* |
| 113: | gb_gss14:* |
| 114: | gb_gss15:* |
| 115: | gb_gss16:* |
| 116: | gb_gss17:* |


```

117:  qb_gss18:*
118:  qb_gss19:*
119:  em_gss13:*
120:  qb_gss20:*
121:  qb_gss21:*
122:  qb_gss22:*
123:  qb_gss23:*
124:  qb_gss24:*
125:  em_gss14:*
126:  em_gss15:*
127:  em_gss16:*

```

| Result No. | Query Score | Match Length | DB ID |
|---------------|----------------|-----------------|----------|
|---------------|----------------|-----------------|----------|

| | | | | | | | |
|---|----|------|-----|------|-----|-----------|---------------------|
| C | 1 | 45.6 | 5.1 | 1101 | 121 | CNS0039G | AL065921 Drosophila |
| C | 2 | 43.4 | 4.8 | 1101 | 121 | CNS00172 | AL078714 Drosophila |
| C | 3 | 40.4 | 4.5 | 1101 | 121 | CNS000B3U | AL055389 Drosophila |
| C | 4 | 39.4 | 4.4 | 552 | 96 | AQ032958J | AQ032958 HS_5001_A |
| C | 5 | 39.2 | 4.4 | 520 | 100 | AQ621653 | AQ621653 HS_3089_J |
| C | 6 | 39 | 4.3 | 750 | 115 | AZ188670 | AZ188670 SP_0170_J |
| C | 7 | 38.8 | 4.3 | 459 | 21 | AM333544 | AM333544 S23N10 AG |
| C | 8 | 38.8 | 4.3 | 515 | 21 | AM333545 | AM333545 S46D3 AGS |
| C | 9 | 38.8 | 4.3 | 545 | 21 | AM334819 | AM334819 S39E11 AG |
| C | 10 | 38 | 4.3 | 250 | 15 | AD060486 | AD060486 AD060486 |
| C | 11 | 38 | 4.2 | 1101 | 121 | CNS000E81 | AL065979 Drosophila |
| C | 12 | 37.6 | 4.2 | 524 | 90 | AQ204415 | AQ204415 HS_3218_A |
| C | 13 | 37.6 | 4.2 | 674 | 121 | CNS00543D | AL320242 Tetradodon |
| C | 14 | 37.6 | 4.2 | 1101 | 121 | CNS000Y4L | AL069827 Drosophila |
| C | 15 | 37.4 | 4.2 | 510 | 97 | AQ413085 | AQ413085 RPCT-11-1 |
| C | 16 | 37.4 | 4.2 | 545 | 98 | AQ534365 | AQ534365 RPCT-11-3 |
| C | 17 | 37.4 | 4.2 | 601 | 96 | AQ375742 | AQ375742 RPCT11-15 |
| C | 18 | 37.4 | 4.2 | 717 | 116 | AZ1969394 | AL1969394 SP_1031_E |
| C | 19 | 37.4 | 4.2 | 1101 | 121 | CNS0106X | AL098595 Drosophila |
| C | 20 | 37.4 | 4.2 | 1101 | 121 | CNS0117XD | AL108619 Drosophila |
| C | 21 | 37 | 4.1 | 510 | 25 | AM943977 | AM943977 LP01628.3 |
| C | 22 | 37 | 4.1 | 938 | 121 | CNS0061J | AL065906 Drosophila |
| C | 23 | 36.8 | 4.1 | 521 | 115 | AZ129098 | AZ129098 OSNB0007 |
| C | 24 | 36.8 | 4.1 | 941 | 123 | CNS0319Y | AL244887 Tetradodon |
| C | 25 | 36.6 | 4.1 | 1101 | 121 | CNS0019R | AL068046 Drosophila |
| C | 26 | 36.2 | 4.0 | 715 | 111 | AZ257353 | AZ257353 RPCT-23-1 |
| C | 27 | 36.2 | 4.0 | 890 | 121 | CNS0179S | AL108346 Drosophila |
| C | 28 | 36.2 | 4.0 | 900 | 36 | C93831 | C93831 C93831 Diet |
| C | 29 | 36.2 | 4.0 | 1100 | 121 | CNS000G3S | AL071398 Drosophila |
| C | 30 | 36 | 4.0 | 391 | 22 | AM472600 | AM472600 s127gm1.y |
| C | 31 | 36 | 4.0 | 533 | 92 | AI166437 | AI166437 xylem.est |
| C | 32 | 36 | 4.0 | 850 | 15 | AD053215 | AD053215 AU053215 |
| C | 33 | 36 | 4.0 | 553 | 122 | CNS023KH | AL179594 Tetradodon |
| C | 34 | 36 | 4.0 | 929 | 121 | CNS016AD | AL106512 Drosophila |
| C | 35 | 36 | 4.0 | 936 | 123 | CNS0465N | AL276764 Tetradodon |
| C | 36 | 36 | 4.0 | 1101 | 121 | CNS001C0 | AL088607 Drosophila |
| C | 37 | 36 | 4.0 | 1101 | 121 | CNS000ZC3 | AL097485 Drosophila |
| C | 38 | 35.8 | 4.0 | 1033 | 121 | CNS0143L | AL103655 Drosophila |
| C | 39 | 35.6 | 4.0 | 1101 | 121 | CNS000LEP | AL088529 Drosophila |
| C | 40 | 35.4 | 3.9 | 1101 | 121 | CNS0000D1 | AL065414 Drosophila |
| C | 41 | 35.4 | 3.9 | 389 | 3 | AA372519 | AA372519 EST84437 |
| C | 42 | 35.4 | 3.9 | 449 | 90 | AQ020742Z | AQ020742Z HS_3218_B |
| C | 43 | 35.4 | 3.9 | 536 | 25 | AM932995 | AM932995 EST358838 |
| C | 44 | 35.4 | 3.9 | 562 | 121 | CNS0160R | AL107013 Drosophila |
| C | 45 | 35.4 | 3.9 | 602 | 25 | AM933405 | AM933405 EST359344 |

ALIGNMENTS

| | | | | | |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-----|-----|-------------|
| LOCUS | CNS00396 | 1101 bp | DNA | GSS | 03-JUN-1999 |
| DEFINITION | Drosophila melanogaster genome survey sequence TEF3 end of BAC # BACR08K10 of RPEI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | | | |
| ACCESSION | AL063921 | | | | |
| VERSION | AL063921.1 GI:494178 | | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | fruit fly. | | | | |
| ORGANISM | Drosophila melanogaster | | | | |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | | |
| REFERENCE | 1 (bases 1 to 1101) Genoscope. | | | | |
| AUTHORS | Direct Submission | | | | |
| TITLE | Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : | | | | |
| JOURNAL | BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr | | | | |
| COMMENT | - Web : www.genoscope.cns.fr determination of this BAC-end sequence was carried out as part of e | | | | |

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osogawa and Aaron Mommser in Piter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

| FEATURES | SOURCE | location/Qualifiers |
|------------|-----------------------|---------------------|
| | 1. 1101 | |
| | /organism="Drosophila | melanogaster" |
| | /dbxref="taxon:7227" | |
| | /clone_1b="RRC1.98" | |
| | /clone="BACRO8K10" | |
| | /note="end : TET3" | |
| BASE COUNT | 201 a | 202 t |
| ORIGIN | 64 c | 131 g |
| | | 503 others |

| | | | | |
|-----------------------|--------|-------------------|-----------------|--------------|
| Query Match | 5.1%; | Score 45.6; | DB 121; | Length 1101; |
| Best Local Similarity | 14.3%; | Pred. No. 0.028; | | |
| Matches | 76; | Conservative 244; | Mismatches 211; | Indels 1; |
| | | | Gaps | 1; |

| | | | |
|----|-------|--------------------------------------------------------------|-----|
| QY | 60 | attgttaataaggagggatttaacatggcgcgcatgtttaagaacccagcaactctgc | 119 |
| Db | 1080. | WTKKMTWMDRRADRRRAGADANMADDGAGTWTATWMTWMTWMTATWTDWMDKMMWAT | 102 |
| QY | 120 | aattgaaggagattgcaaatccggtcttaatacgcgtccgattgtgttaactatgaggg | 179 |
| Db | 1020 | AAKTDATWTTWMTATWRADWAGRGDRGAKRDRDAATADAGRGDRGKRKKDKDRKGDD | 961 |
| QY | 180 | acaaatgacaaaagatgacatccatccagttaagaacactatctctttgcggaagata | 239 |
| Db | 960 | DKRGKKKKAAKAAKATWTKMDWMDWMDKMKMGAGADRADDGAGDKDDGCKRADDD | 901 |
| QY | 240 | tcaattgtttgcgtcttgtaagttcatgattacaggttatgtatcatcttgcttcgct | 299 |
| Db | 900 | DYDTCDDDDDKMDWMDKKGTTWGSATWMAATDMMWGMADMDWMTDAADMDWAD | 841 |
| QY | 300 | caatcgtgcgttgatattgattgataatgaagaagtgtcttaattatggaagaagatcc | 359 |
| Db | 840 | RMDAMAKMDDAAMGAKRTADRRDMDRAGKRGAKRRDRKRAADKRDAADRRDAAW | 781 |
| QY | 360 | cgtcatattatcaatctgcgaatagtgattgtgttcgttgcggaaggagtgcttggtcga | 419 |
| Db | 780 | TTWTTTTTDTDDMKKTDPTWEMADPRWMDRDDDDDRBAGTAGCKMKRTTKRRKKRD | 721 |
| QY | 420 | cgggtataaacaag-caatcccgcgattgcgtaacgcggtcttaaacatcaacttgatg | 478 |

| DB | 720 | TRMDDADADDTTARRRRRGGADGAGKCKTKGRKRRRRDRATWDRDAMWDAAMWTTT | 661 |
|------------|-----|------------------------------------------------------------|-----|
| 0y | 479 | tgaagctgcgggggtggggagcaatttcccaatcgattcatatgatagaagaagattt | 538 |
| Db | 660 | DDTDDDMKRRRRRRGARRRRRTTAAADMTWTWAKMDAKMDKTRADRWDRMAADTWT | 601 |
| 0y | 539 | ttaagctgaccctacaagaataacatggttcattgatcatatgtatgata | 590 |
| Db | 600 | DARKRDRMAKARARARDRARARADRWTTTKGTTATTATTAAARAW | 549 |
| RESULT | 2 | | |
| CNS00LT2 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |
| FEATURES | | | |
| BASE COUNT | | | |
| ORIGIN | | | |

[illegible]

```

Db 497 AMCTCTTAKKKKKKKGGTAAGAKKAKKAAAMTKATKTKKAAAMADTATTA 438
OY 365 ttataatattgcgaatgaagtgttgcttcgaggaaggagatgcttgagcgaaggt 424
Db 437 KAKKKAATTAADGATGTGTTKTKAKTKAGKRGDGAAMAAATGKTATKAGK 378

RESULT 4
AO332958 552 bp DNA GSS 06-MAR-1999
LOCUS HS.5001.A2.F09.T7.RPc11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=577 Col=18 Row=K, DNA sequence.
ACCESSION AO332958
VERSION AO332958.1 GI:4130585
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 577 row: K column: 18
Seq primer: T7
Class: BAC ends
High quality sequence stop: 552.
Location/Qualifiers
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate="577 Col=18 Row=K"
/clone_lib="RPc11 Human Male BAC Library"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; RPc11 Human Male BAC Library"

BASE COUNT 177 a 133 c 67 g 170 t 5 others
ORIGIN

Query Match 4.4%; Score 39.4; DB 96; Length 552;
Best Local Similarity 53.7%; Pred. NO. 1.3;
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

OY 204 taccagtaagaacattatcttccttagcggaagataatcattgtgtgtctgaagt 263
Db 182 TAAAGTAGAATATCTATTAAGTAGAGTAAGAAATATATCTTGCTCTGATGATGAT 123

OY 264 tcaatgacgtacccgttatgattcattccttcgttcgttcgaatgctgtgtatattgat 323
Db 122 CCAAAACAGCAGCTTTTGAGTTTGAAGCAGATCATTCAATGTTCTATGAAATTTGAGT 63

OY 324 tgaatgagaagtgcttcaattggaag 350
Db 62 TGATGAGATAGTAATCAATCGAGAGA 36

RESULT 5
AO621653

```

```

LOCUS AO621653 520 bp DNA GSS 16-JUN-1999
DEFINITION HS.3089_B2.H12.MF CTR Approved Human Genomic Sperm Library D Homo
ACCESSION sapiens genomic clone Plate=3089 Col=24 Row=P, DNA sequence.
VERSION AO621653
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 520)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3089 row: P column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 520.
Location/Qualifiers
1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate="Plate=3089 Col=24 Row=P"
/clone_lib="CTR Approved Human Genomic Sperm Library D"
/sex="Male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 146 a 100 c 129 g 143 t 2 others
ORIGIN

Query Match 4.4%; Score 39.2; DB 100; Length 520;
Best Local Similarity 57.3%; Pred. NO. 1.5;
Matches 71; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 223 tcttagcggaagataatcattgtgtgttccttgaaagtctatgctacccggttat 262
Db 211 TCTTCTCTCAATACAGACTTCTTAATTTTCTGTTAAGTTCCTCAGGTTT 270

OY 283 gattccattgtctgcctcaatcgcgtgtgtattgattgaatgaatgaagatgctta 342
Db 271 GTTAAATTTGATGATGTTTCAGATTTTCGATTCAGATTCGCAACTAAAGAGATGCAATTTA 330

OY 343 attg 346
Db 331 AATG 334

RESULT 6
AO138670
LOCUS AO138670 750 bp DNA GSS 06-JUN-2000
DEFINITION SP.0170.A2.C04.SP6E Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=170 Col=8 Row=E, DNA sequence.
ACCESSION AO138670
VERSION AO138670.1 GI:8290573
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

```


REFERENCE
AUTHORS
Roest Croollius,H., Jalllon,O., Dasilva,C., Ozouf-Costaz,C., Filames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,M., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10, 939-949 (2000)

TITLE
Tetraodontidae, Tetraodon.

JOURNAL REFERENCE
AUTHORS
1 (bases 1 to 674)
2 (bases 1 to 674)
Roest Croollius,H., Jalllon,O., Dasilva,C., Ozouf-Costaz,C., Bernot,A., Filames,C., Winkler,P., Brotlier,P., Quetier,F., Saurin,M. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
Nat.Genet. 25 (2), 235-238 (2000)

MEDLINE
JOURNAL TITLE
AUTHORS
3 (bases 1 to 674)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/genbank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
Location/Qualifiers
1..674
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="067120"
/clone_1lb="C"
/note="Genoscoope sequence ID : C0CC067BE10CI-end : T7"

BASE COUNT
ORIGIN
168 a 113 c 165 g 192 t 36 others

Query Match
Best Local Similarity 4.2% Score 37.6; DB 124; Length 674;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Dy 472 ttgaagttagatgcctcgccggggtcggagacattccaccatcatcattgatgaaaga 531
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 TTTATTTCNGATNCNMGNGGTGCTCCANGTTCACCAAGTGATGCACATTTGTATTAG 566
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 532 gaagttttaatgctgcacctcaacgaataacaatgatttcgattcataatgaatat 591
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 NAAAGCGAAGTATCGAAGCTTAACAACGTGCTTTTTTTCGNMTCAAAGATGAATTT 626
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Dy 592 gcaggtgtaatgcatgcgaattcgtactaatatt 627
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 TCACAGCTTAATAACTACCAAATTTAAAGAANAAGTT 662
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14

CNS00YWL 1101 bp DNA GSS 26-JUL-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence SpE end of BAC
BACN01G13 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL096927
AL096927 .1 GI:5608538
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Genoscoope.
Direct Submission
Submitted (23-JUL-1999) Genoscoope - Centre National de Sequencage :

| | | | | | |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|-------------|---------|------------|
| | | | | | |
| COMMENT | BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..1101 /oranism="Drosophila melanogaster" /plasmid="pBelobAC11" /db_xref="taxon:7227" /clone_1lb="DrosBAC" /clone="BACN01G13" /note="end : SP6" | | | | |
| BASE COUNT | 424 a | 119 c | 129 g | 248 t | 181 others |
| ORIGIN | | | | | |
| Query Match | 4.2%; Score 37.6; DB 121; Length 1101; | | | | |
| Best Local Similarity | 30.5%; Pred. No. 5.1; | | | | |
| Matches | 99; Conservative | 62; Mismatches | 164; Indels | 0; Gaps | 0; |
| OY | 323 ttgaaatgagaagtgctttaatgtgaagaagaataccgcgttaattaatatgtcgaatg | 382 | | | |
| Dd | 630 TTACAAGAAGANGTGTTCCTGCCTTCYCCKMCCCCCCTTAGGABATGKCGCGTA | 689 | | | |
| OY | 383 aatggttcgttcgtgggaaggaggtctgttggcgcgagagtataacaagaatccgcg | 442 | | | |
| Dd | 690 TTTGGGGRTRMKGTGGGTAAATKTGGGGGKGKGGGGGMATTARARRGGAAGRACCT | 749 | | | |
| OY | 443 gatcgcgaacgccgcgtctaacaacataacctgatgtatgatatgctcgggggtgggacaat | 502 | | | |
| Dd | 750 KGTGGRTGKAAMGCCMTAMAVSSBBKKKKKKRGKRVTGKMTTGTTTTTTANAA | 809 | | | |
| OY | 503 ttccaacatcgattcatgattatggaagaagaagttttaatgctgcgaaccctcaagaata | 562 | | | |
| Dd | 810 ATWAAAAAATAATATNMWATAAAAATAWAAARAANNATWTATATAAAMWMAAAATATA | 869 | | | |
| OY | 563 caatgttcgatcataatgfatgaatltggaagtygtaatgcatcgcaagtlcglaacta | 622 | | | |
| Dd | 870 ATAAATTAATAAATAAATAAATAATATATATATWAAAAAAATAATATTAAMAAATWAW | 929 | | | |
| OY | 623 atatgaccagatctctaatacaaga | 647 | | | |
| Dd | 930 AKATAAAMWMAAAAMWMAATWAMA | 954 | | | |
| RESULT 15 | | | | | |
| AQ413085 | GSS 23-MAR-1999 | | | | |
| LOCUS | AQ413085 510 bp DNA | | | | |
| DEFINITION | RPCR-11-170D10.TJ RPCR-11 Homo sapiens genomic clone RPCR-11-170D10 | | | | |
| ACCESSION | 'DNA sequence.' | | | | |
| VERSION | AQ413085 | | | | |
| KEYWORDS | AQ413085.1 GI:4472573 | | | | |
| SOURCE | GSS. | | | | |
| ORGANISM | human. | | | | |
| | Homo sapiens | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 510) | | | | |
| REFERENCE | Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter | | | | |
| AUTHORS | J.C. | | | | |
| TITLE | Use of BAC End Sequences from Library RPCR-11 for Sequence-Ready | | | | |
| JOURNAL | Map Building | | | | |
| COMMENT | Unpublished (1997) Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 | | | | |

us-09-339-1_copy_91_990.rst

... For BAC
used from
ing) or from
age:
_search.html.

2: ECOR1;

| | |
|----|------|
| 0; | 0; |
| 0; | Gaps |
| 0; | 0; |

ttcaat 303
|||
TTTCAG 260

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:29:50 ; Search time 214.78 Seconds
(without alignments)
78.010 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNGFKRFFSTLSTLLASS.....QSASDSGQSTSYIDNVAIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 2607 | 100.0 | 490 | 21 | Y54122 |
| 2 | 1804 | 69.2 | 493 | 21 | Y44495 |
| 3 | 1804 | 69.2 | 493 | 21 | Y54124 |
| 4 | 1711 | 65.6 | 468 | 21 | Y44496 |
| 5 | 1711 | 65.6 | 468 | 21 | Y54125 |
| 6 | 1640 | 62.9 | 476 | 21 | Y54123 |
| 7 | 1152 | 44.2 | 331 | 21 | Y54127 |
| 8 | 1075 | 41.2 | 369 | 21 | Y54129 |
| 9 | 957.5 | 36.7 | 320 | 21 | Y54132 |
| 10 | 750.5 | 28.8 | 305 | 21 | Y54130 |
| 11 | 546 | 20.9 | 188 | 21 | Y54133 |
| 12 | 419 | 16.1 | 487 | 10 | P91007 |

| | | | | | |
|----|-------|------|------|----|--------|
| 13 | 318 | 12.2 | 132 | 21 | Y54131 |
| 14 | 229.5 | 8.8 | 400 | 18 | W23601 |
| 15 | 229.5 | 8.8 | 400 | 18 | W22521 |
| 16 | 229.5 | 8.8 | 400 | 19 | W57431 |
| 17 | 229.5 | 8.8 | 462 | 19 | W57433 |
| 18 | 226.5 | 8.7 | 410 | 18 | W12378 |
| 19 | 226 | 8.7 | 411 | 18 | W12379 |
| 20 | 224 | 8.6 | 409 | 9 | P81843 |
| 21 | 214 | 8.2 | 1010 | 19 | W34989 |
| 22 | 213 | 8.2 | 467 | 17 | W05731 |
| 23 | 213 | 8.2 | 467 | 17 | W00382 |
| 24 | 211 | 8.1 | 411 | 18 | W12381 |
| 25 | 209.5 | 8.0 | 412 | 18 | W12380 |
| 26 | 193 | 7.4 | 499 | 14 | R42122 |
| 27 | 189 | 7.2 | 351 | 18 | W18790 |
| 28 | 186.5 | 7.2 | 352 | 20 | W08472 |
| 29 | 179.5 | 6.9 | 476 | 19 | W37241 |
| 30 | 175 | 6.7 | 484 | 19 | W43910 |
| 31 | 169 | 6.5 | 832 | 13 | R25021 |
| 32 | 166 | 6.4 | 941 | 11 | R07478 |
| 33 | 166 | 6.4 | 941 | 11 | R77395 |
| 34 | 164 | 6.3 | 800 | 8 | P70420 |
| 35 | 160 | 6.1 | 531 | 16 | W01503 |
| 36 | 158.5 | 6.1 | 472 | 19 | W37243 |
| 37 | 154.5 | 5.9 | 532 | 12 | R13229 |
| 38 | 151.5 | 5.8 | 562 | 21 | Y65008 |
| 39 | 150 | 5.8 | 357 | 16 | R77394 |
| 40 | 143 | 5.5 | 521 | 17 | R89927 |
| 41 | 142.5 | 5.5 | 956 | 19 | W49874 |
| 42 | 142 | 5.4 | 358 | 19 | W39262 |
| 43 | 129 | 4.9 | 1536 | 14 | R41723 |
| 44 | 129 | 4.9 | 1536 | 14 | R41725 |
| 45 | 129 | 4.9 | 1536 | 15 | R63505 |

ALIGNMENTS

| | |
|----------|--------------------------------------------------------------------------|
| RESULT 1 | |
| Y54122 | Y54122 standard; Protein: 490 AA. |
| XX | XX |
| AC | Y54122; |
| DT | 27-MAR-2000 (first entry) |
| XX | XX |
| DE | Amino acid sequence of a Bacillus mannanase enzyme. |
| XX | XX |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; |
| KW | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; glucomannan; |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; yam; fabric; |
| KW | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; |
| KW | printing paste; plant material degradation; recycled waste paper; |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; |
| KW | mannan-containing food; coffee extract; cleaning composition; |
| KW | machine washing; hard-surface cleaner; dishwashing; oral; dental; |
| KW | contact lens; body-care composition; fabric softener; oil well drilling; |
| KW | subterranean formation fracture. |
| XX | XX |
| OS | Bacillus sp. |
| XX | XX |
| FH | Key |
| FT | Peptide |
| FT | 1..30 |
| FT | /note- "signal peptide" |
| FT | Domain |
| FT | 31..330 |
| FT | /note- "catalytic domain" |
| FT | Misc-difference |
| FT | 116 |
| FT | /note- "encoded by CTR" |
| FT | Domain |
| FT | 331..342 |
| FT | /note- "linker" |
| FT | 343..490 |
| FT | /note- "domain of unknown function" |
| XX | XX |

Amino acid sequence
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
P300-CeLB fusion c
P300-CeLB fusion c
Sequence of alkali
Teredinbacter end
Cellulase. Bacill
Bacillus cellulase
P300-CeLB fusion c
P300-CeLB fusion c
NK-1 cellulase. B
Corrected Bacillus
Actinomyces sp. 3
Heterodera glycine
Heterodera glycine
Alkaline cellulase
Cellulase. Bacill
Full length Bacill
Sequence encoded b
60 kD endoglucanase
Globodera rostoch
Endoglucanase enco
Acidothermus cellu
Fragment of alkali
A. cellulolyticus
Banksia gouldi glyc
A. cellulolyticus
High molecular wei
High molecular wei
Haemophilus high m


```
XX
SQ Sequence 493 AA;
Query Match 69.2%; Score 1804; DB 21; Length 493;
Best Local Similarity 69.2%; Pred. No. 1.2e-126;
Matches 332; Conservative 64; Mismatches 80; Indels 4; Gaps 3;
OY 11 ITLSLLASSILFVSGSTANANGFYVSGTLLDANGNPFVMGINGHAMYKDOATTA 70
D 12 IICLLISVGMGLT-TPSAASATGYVDGNTLYDANGQPFVMTGINGHAMYKDTASTA 70
OY 71 IEGANGANTVRIYLSGGGOWTKODIHTVNLISLAEEDNLVAVPEYHATGDSIASL 130
D 71 IpaLaegqanfirivlsdgqwekdldltrevlelaeqnmvavevhdatsrdsd 130
OY 131 NRAVDYWMRSALIGKEDVTIINANEMFGSWGDAMADYKQAIPLRANAGLHNTLMV 190
D 131 nrvdywlemkdallgkedvtlinanewygsawsawadyidvlpkldagltlclmv 190
OY 191 DAAGWGFPOSIDHYGREVNADPQRTMFSIHMYETAGNASOVRNIDRYLNDIALV 250
D 191 daagwgyppgsihdygqdvfnadplkntmfshmyeyagdantrsnldtrvldqdlaly 250
OY 251 IGEGRHHTNGDVDEATIMSYSEQRGVGKLAWSKNGKPEWEYIDLSDMGNNTANGN 310
D 251 Igegrhhtdgdvdedltllyseeqyglawswkngstewdyldlseedwagqhltdvgn 310
OY 311 TIVNGPYLRSTSLVFT--GGSGDGTSPPTLYDEFGSMOGWTSLSGGPVAWTEW 368
D 311 rivngadglqetstkrstvtfdiddngshpeptatcltydfeqstgvygshvngpysvlew 370
OY 369 SSKSHSIKADIQSSNQHTLVHVIQNTSLQNSRIQATVHAMGVSNGMTARLYKT 428
D 371 gaagyslkadvnltsnshelysegrnllyeqqlnatvrhanvgnpgnqmarlykt 430
OY 429 GHGTWVSGSVPIGNS-SGTTLSIDLSNVQNTLSQVREIGVQFOSASSSGOTSITDNY 487
D 431 gsdylwmspgrltnsnsgtllsfidlnhshvirelygqfsadnssqtalvadvn 490
RESULT 3
ID Y54124 standard; Protein; 493 AA.
XX
AC Y54124;
XX
DT 27-MAR-2000 (first entry)
XX
DE Amino acid sequence of a Bacillus mannanase enzyme.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture.
XX
OS Bacillus agaradhaerens.
XX
XX Key Location/Qualifiers
XX Peptide 1..31
XX FT /note- "signal peptide"
XX FT Domain 32..344
XX FT /note- "catalytic domain"
XX FT Domain 345..493
XX /note- "domain of unknown function"
XX
XX MO9964619-A2.
XX
XX
```

```
PD 16-DEC-1999.
XX
XX 10-JUN-1999; 99WO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
XX 20-OCT-1998; 98DK-0001340.
XX 20-OCT-1998; 98DK-0001341.
XX 28-OCT-1998; 98US-0105970.
XX 28-OCT-1998; 98US-0106054.
XX 23-DEC-1998; 98DK-0001725.
XX 05-MAR-1999; 99DK-0000306.
XX 05-MAR-1999; 99DK-0000307.
XX 05-MAR-1999; 99DK-0000308.
XX 05-MAR-1999; 99DK-0000309.
XX 09-MAR-1999; 99US-0123543.
XX 10-MAR-1999; 99US-0123623.
XX 11-MAR-1999; 99US-0123641.
XX 11-MAR-1999; 99US-0123642.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX N-PSDB: Z45337.
XX
XX New mannanases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions
XX
XX Claim 35; Page 213-214; 242pp: English.
XX
XX The present sequence represents a mannanase (also known as mannan
XX endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
XX mannanase hydrolyses galactomannans. Specifically, mannanases hydrolyse
XX 1,4-beta-D-mannosidic linkages in mannan, galactomannans, glucomannans,
XX and galactoglucomannans. The mannanase protein, or preparations
XX containing it, are used to improve properties of cellulosic or
XX synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannanases
XX are also used to process coffee extracts (to inhibit gel formation);
XX in cleaning compositions (for machine washing of fabrics, as
XX hard-surface cleaners, for hand or machine dishwashing, also in oral,
XX dental, contact lens or body-care compositions) where they remove
XX mannan-containing soils and prevent binding of some soils to
XX cellulosics; and in fabric softeners. They can also be used in oil
XX well drilling to fracture subterranean formations.
XX
XX
XX SQ Sequence 493 AA;
Query Match 69.2%; Score 1804; DB 21; Length 493;
Best Local Similarity 69.2%; Pred. No. 1.2e-126;
Matches 332; Conservative 64; Mismatches 80; Indels 4; Gaps 3;
OY 11 ITLSLLASSILFVSGSTANANGFYVSGTLLDANGNPFVMGINGHAMYKDOATTA 70
D 12 IICLLISVGMGLT-TPSAASATGYVDGNTLYDANGQPFVMTGINGHAMYKDTASTA 70
OY 71 IEGANGANTVRIYLSGGGOWTKODIHTVNLISLAEEDNLVAVPEYHATGDSIASL 130
D 71 IpaLaegqanfirivlsdgqwekdldltrevlelaeqnmvavevhdatsrdsd 130
OY 131 NRAVDYWMRSALIGKEDVTIINANEMFGSWGDAMADYKQAIPLRANAGLHNTLMV 190
D 131 nrvdywlemkdallgkedvtlinanewygsawsawadyidvlpkldagltlclmv 190
OY 191 DAAGWGFPOSIDHYGREVNADPQRTMFSIHMYETAGNASOVRNIDRYLNDIALV 250
D 191 daagwgyppgsihdygqdvfnadplkntmfshmyeyagdantrsnldtrvldqdlaly 250
```

```

OY 251 IGEFGRHRTNGDVDEATIMSYSEORGVGLAMSKNGPMEYELDLSDNAGNNTLTMGN 310
DB 251 Igefhrtngdvdedtlisyeetgvlawskngstewdyldlbedagqhltdwgn 310
OY 311 TIVNGPYGLRETSRLSTVFT--GGGSDGGSPTLLYDFEGSMQGTSSLSGGPAAVTEW 368
DB 311 rlvbgadglgetskpstvftddngghpeptatcllydfegstgwhsnvsgpawvlew 370
OY 369 SSKGSHSLKADLOLSNSQHTLVIONTSIQONSRIQATVHANWGSYNGMTARLYVKT 428
DB 371 gasgnyalkadvnltsnshelyeegtrnlhysqinatvhanvgnpgnmarllyvkt 430
OY 429 GHGYTWYSGSFVPINGS-SGTTLSIDLNSVQNLSCVREIGVQFSASDSSGTSITIDNV 487
DB 431 gsdylwshgpftrlnssngstltsfdlnlnshvrelvgvfasadnsqgtalydnhv 490

RESULT 4
Y44496
ID Y44496 standard; Protein; 468 AA.
AC Y44496;
DE 27-MAR-2000 (first entry)
XX Bacillus agaradherens Clone MB594, Mannanase enzyme.
XX
XX Mannanase: endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
XX detergent composition: mid-branched anionic surfactant; washing;
XX cosmetic stain; food stain.
XX
XX Bacillus agaradherens.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Peptide 1..31
XX Protein /label- Signal_peptide
XX 32..468
XX /label- Mature_Mannanase
XX
XX W09964552-A1.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Bettiol JP, Thoen CAJK;
XX
XX WPI; 2000-116536/10.
XX
XX N-PSDB; 229846.
XX
XX Detergent composition for removing greasy stains such as cosmetics,
XX food stains and body soils
XX
XX Disclosure; Page 103-104; 113pp; English.
XX
XX The present sequence is the B. agaradherens Clone MB594, alkaline
XX Mannanase enzyme. Clone MB594 is derived from B. agaradherens strain
XX NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
XX from 7.5-10.5. It can be used in a detergent composition along with a
XX mid-branched anionic surfactant. The detergent composition may be used
XX for washing purposes, soaking/pretreatment of stained fabric, hard
XX surface cleaning and for removal of cosmetic and/or food stains. This
XX composition provides excellent cleaning effect at low temperature.
XX
XX Sequence 468 AA;

```

```

Best Local Similarity 69.3%; Pred. No. 9,2e-120;
Matches 314; Conservative 60; Mismatches 75; Indels 4; Gaps 3;

OY 11 ITSLILASSILFVSGSTPYANNSGFFVSGTLLYDANGNPVFMGGINGHANYKQOATTA 70
DB 12 lltcltlisvnglgt-tcpsaastgfyvngncllydangqpfvmrglnhghavkdtasta 70
OY 71 IESANTGANTVAVIVSDGGQRTKDIHVRNLISLAEDNHLVAAPEVADATGYOSIASL 130
DB 71 lpaiaegantllrlivlsdggqekddltirevlaeqnmvavevndacgrdsrdl 130
OY 131 NRVADYIEMRSALICEPEVIVINIANEMFSGWEGDAMADGYKQALPRLRNGLNHTLV 190
DB 131 nrvadvylemkdallkbedvllnlanewysgdswadgyldvlpklrdaqltcllmv 190
OY 191 DAAGWGFQPSIHIDYREVFNADPQRNMFSTIMYETIAGNSQVATNIDRLNOLALV 250
DB 191 daagwgfypslndygdvfnadpqlkntmflmyeyaggdantvrsndryldqdlav 250
OY 251 IGEFGRHRTNGDVDEATIMSYSEORGVGLAMSKNGPMEYELDLSDNAGNNTLTMGN 310
DB 251 Igefhrtngdvdedtlisyeetgvlawskngstewdyldlbedagqhltdwgn 310
OY 311 TIVNGPYGLRETSRLSTVFT--GGGSDGGSPTLLYDFEGSMQGTSSLSGGPAAVTEW 368
DB 311 rlvbgadglgetskpstvftddngghpeptatcllydfegstgwhsnvsgpawvlew 370
OY 369 SSKGSHSLKADLOLSNSQHTLVIONTSIQONSRIQATVHANWGSYNGMTARLYVKT 428
DB 371 gasgnyalkadvnltsnshelyeegtrnlhysqinatvhanvgnpgnmarllyvkt 430
OY 429 GHGYTWYSGSFVPINGS-SGTTLSIDLNSVQNLSCVREIGVQFSASDSSGTSITIDNV 487
DB 431 gsdylwshgpftrlnssngstltsfdlnlnshvrelvgvfasadnsqgtalydnhv 490

RESULT 5
Y54125
ID Y54125 standard; Protein; 468 AA.
AC Y54125;
DE 27-MAR-2000 (first entry)
XX
XX Amino acid sequence of a Bacillus mannanase enzyme.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX mannan; galactomannan; 1,4-beta-D-mannosidic linkage;
XX mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
XX synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
XX plant material degradation; recycled waste paper; paper making pulp;
XX guar; locust bean gum; mannan-containing food; coffee extract;
XX cleaning composition; machine washing; hard-surface cleaner;
XX dishwashing; oral; dental; contact lens; body-care composition;
XX fabric softener; oil well drilling; subterranean formation fracture.
XX
XX Synthetic.
XX Bacillus agaradherens.
XX
XX W09964619-A2.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1999; 99WO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
XX
XX 20-OCT-1998; 98DK-0001340.
XX
XX 20-OCT-1998; 98DK-0001341.
XX
XX 28-OCT-1998; 98US-0105970.
XX
XX 28-OCT-1998; 98US-0106054.
XX
XX 23-DEC-1998; 98DK-0001725.
XX
XX 05-MAR-1999; 99DK-0000306.
XX
XX 05-MAR-1999; 99DK-0000307.

```

PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99DS-0123543.
PR 10-MAR-1999; 99DS-0123623.
PR 10-MAR-1999; 99DS-0123641.
PR 11-MAR-1999; 99DS-0123642.
PA (NOVO) NOVO-NORDISK AS.
PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PT WPI: 2000-105891/09.
PS N-PSDB: Z45338.
XX
XX New mannases for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions -
XX
XX Example 5; Page 215-216; 242pp: English.
XX
XX The present sequence represents a mannanase (also known as mannan
CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The present mannanase is a synthetic variant of the mannanase of
CC Y54124, in which the C-terminus of the protein was changed due to
CC design of a lower PCR primer used for amplification. The mannanase
CC hydrolyses galactomannans. Specifically, mannanases hydrolyse
CC 1,4-beta-D-mannosidic linkages in mannan, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulostic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC celluloses; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.
XX
XX Sequence 468 AA:
SQ

```

Query Match          65.6%; Score 1711; DB 21; Length 468:
Best Local Similarity   69.3%; Pred. No. 9,2e+120:
Matches    314; Conservative    60; Mismatches    75; Indels     4; Gaps      3

QY 11 ITLLSLASSILFEVSGTSTANANSGEFYSGITLYDANGNPFYMRGINHGHWKDOATTA 70
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 12 lictllsavgmglr-tapsasstgfydgnglilydangprfymrglnbhaykclaststa 70
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 71 IEGIANGTANYRIVLYSDGOGMTKDDIHTVRNLISLAEDNHLVAYPEVDHATGYDSIASL 130
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 71 lpalaegagantlrilysdgggwekddidlirevialeeqnkmavavevdhatcgrdsrl 130
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 131 NRAVDYTMEMRSALIGKETVTIINTANEMFGSWEDGANWADYRKAIIPRRMAGNHTLMV 190
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 131 nraavytlemkdallqkedvtvlinaanevsgwsdgsawadgyldvlpkrldagltclmrv 190
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 191 DAAGNGEPOSIIHDYGREVFNAIDPORNTWFSTIHMEYAGAGNSQVRTNIDRYLNODIALY 250
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 191 daagngyqpssihdygdgtfinadplknmfslhmeyegsdanrtvrsndrryldcdtlaly 250
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 251 IGESHRHTNGDYDEATTIKSYSEQRGVGLANSWTGNGPEWEYELDLISNDAGNNITLAWGN 310
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 251 igeshrhnttdcvdedtilisyeecgtgylawsxkgnsctewdyldlsedwagghltdwgn 310
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 311 TVNVPPYGRBTSRLSTVFT--GGSDGETSPITLYDFEGSNMGOWTGSSLGGPNAVTFEM 368
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 311 rlvnvaqdlqeakspscvftclddngnhpepptcltyldiegsqgvwngrvcgspsvetw 370
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 369 SSKGSHSLKADIQLSSNQHYLHVIONTSILOQNSRIQAIVVKRANNGSVGNGMTARLYVT 428

```

| | | | |
|--------|--------------------------------------------------------------------------|------------------------------------------------------------|-----|
| Db | 371 | gasgnylskadvnltstnsbshelyegsrnllygsqlnatvrhanwmpngmarlyvkt | 430 |
| Oy | 429 | GHGWTWYSGSFVPINGS-SGTTLSUDLSNVOL | 460 |
| Db | 431 | gsdylwmspflrlnsmnsqtltsfdlnnlenl | 463 |
| RESULT | 6 | | |
| ID | Y54123 | standard; Protein; 476 AA. | |
| XX | Y54123; | | |
| XX | 27-MAR-2000 | (first entry) | |
| DE | A mannanase-linker-cellulose binding domain fusion protein. | | |
| XX | | | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | | |
| KW | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; | | |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; | | |
| KW | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; | | |
| KW | printing paste; plant material degradation; recycled waste paper; | | |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; | | |
| KW | mannan-containing food; coffee extract; cleaning composition; | | |
| KW | machine washing; hard-surface cleaner; dishwashing; oral; dental; | | |
| KW | contact lens; body-care composition; fabric softener; oil well drilling | | |
| KW | subterranean formation fracture; cellulose binding domain. | | |
| XX | | | |
| OS | Synthetic. | | |
| OS | Bacillus sp. | | |
| OS | Clostridium thermocellum. | | |
| XX | | | |
| XX | WO9964619-A2. | | |
| XX | | | |
| PD | 16-DEC-1999. | | |
| XX | | | |
| PF | 10-JUN-1999; | 99WO-DK00314. | |
| XX | | | |
| PR | 10-JUN-1998; | 98US-0111256. | |
| PR | 20-OCT-1998; | 98DK-0001340. | |
| PR | 20-OCT-1998; | 98DK-0001341. | |
| PR | 28-OCT-1998; | 98US-0105970. | |
| PR | 28-OCT-1998; | 98US-0106054. | |
| PR | 23-DEC-1998; | 98DK-0001725. | |
| PR | 05-MAR-1999; | 99DK-0000306. | |
| PR | 05-MAR-1999; | 99DK-0000307. | |
| PR | 05-MAR-1999; | 99DK-0000308. | |
| PR | 05-MAR-1999; | 99DK-0000309. | |
| PR | 09-MAR-1999; | 99US-0123543. | |
| PR | 10-MAR-1999; | 99US-0123623. | |
| PR | 10-MAR-1999; | 99US-0123641. | |
| PR | 11-MAR-1999; | 99US-0123642. | |
| PA | (NOVO) NOVO-NORDISK AS. | | |
| XX | | | |
| PI | Kauppinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME; | | |
| XX | WPI; 2000-105891/09. | | |
| DR | N-PSDB; 245336. | | |
| XX | | | |
| PT | New mannanases for treatment of textiles, plant material and coffee | | |
| PT | extract, and in cleaning compositions | | |
| XX | | | |
| PS | Example 4; Page 211-212; 242pp; English. | | |
| CC | The present sequence represents a mannanase-linker-cellulose binding | | |
| CC | domain fusion protein. Mannanase (also known as mannan | | |
| CC | endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase) | | |
| CC | hydrolyses galactomannans. Specifically, mannanases hydrolyse | | |
| CC | 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, | | |
| CC | and galactoglucomannans. The mannanase protein, or preparations | | |
| CC | containing it, are used to improve properties of cellulosic or | | |

| | | | | |
|--------|----|---------------|-------------------------------------------------------------------------|-----|
| Dd | | 11 | alicslllfagvlnltssgaeahngsfhvngttcllydangnprfcmrflnghgwfkelet | 70 |
| Oy | | 70 | AIEGIANTGANTVRIIVSDGGQTKDDIHYRNLIISLAEDNHLVAPEVDATGIDSIAS | 129 |
| Dd | | 71 | smrgisqantllivlinsnggrwqkddrmnvaaaislaeqhgmrlavlevhdatsnfad | 130 |
| Oy | | 130 | LNRADVIEIRMSALLICKEDFVIININEMFGSMEGDAMADGKOAIPRLRNGLNHTLM | 189 |
| Dd | | 131 | lgaadvylemkdvdlqgvedvlvlininegyawgdggawerqgnalirqlrnaqslhtlm | 190 |
| Oy | | 190 | VDAAGWGFPOSIDHGREYFNADPQRNTFFSIIHMEYAGGNASQVRINIDRYLNDOLAL | 249 |
| Dd | | 191 | vdaegygsypsvvdyyqgevlinaepqrntfnfsrhmyeyagdeantrrrndslsqnlal | 250 |
| Oy | | 250 | VIEFGHRHTNGVDDEATIMSYSEQRGVGLAWSMKWGKSGEWETLDLSNDMAGNNLTAWG | 309 |
| Dd | | 251 | vifegfhwydgddedtlilsqgrnvglawswhgneqsegveylldlsndfagnrltwyg | 310 |
| Oy | | 310 | NTIVNGPYGLAETSRLSTVF | 329 |
| Dd | | 311 | drlvnpngrlrqtksrvsvf | 330 |
| RESULT | | 8 | | |
| VS4129 | | VS4129 | standard; Protein; 369 AA. | |
| XX | AC | VS4129; | | |
| Dt | XX | 27-MAR-2000 | (first entry) | |
| XX | | | Amino acid sequence of a Bacillus sp. A349 mannanase enzyme. | |
| XX | | | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | |
| KW | | | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | |
| KW | | | mannan; galactomannan; glucomannan; galactoglucomannan; cellulose fibre | |
| KW | | | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | |
| KW | | | plant material degradation; recycled waste paper; paper making pulp; | |
| KW | | | guar; locust bean gum; mannan-containing food; coffee extract; | |
| KW | | | cleaning composition; machine washing; hard-surface cleaner; | |
| KW | | | dishwashing; oral; dental; contact lens; body-care composition; | |
| KX | | | fabric softener; oil well drilling; subterranean formation fracture. | |
| OS | | | Bacillus sp. | |
| XX | | | | |
| XX | FH | Key | Location/Qualifiers | |
| FT | FT | Peptide | 1..25 | |
| FT | | | /note= "signal peptide" | |
| FT | | Domain | 26..67 | |
| FT | | | /note= "N-terminal domain of unknown function" | |
| FT | | Domain | 68..369 | |
| FT | | | /note= "catalytic domain" | |
| XX | | | | |
| PN | | MO9964619-AZ. | | |
| PD | | 16-DEC-1999. | | |
| XX | PF | 10-JUN-1999; | 99MO-DK00314. | |
| XX | | | | |
| PR | | 10-JUN-1998; | 98US-0111256. | |
| PR | | 20-OCT-1998; | 98DK-0001340. | |
| PR | | 20-OCT-1998; | 98DK-0001341. | |
| PR | | 28-OCT-1998; | 98US-0105970. | |
| PR | | 28-OCT-1998; | 98US-0106054. | |
| PR | | 23-DEC-1998; | 98DK-0001175. | |
| PR | | 05-MAR-1999; | 98DK-0000306. | |
| PR | | 05-MAR-1999; | 98DK-0000307. | |
| PR | | 05-MAR-1999; | 99DK-0000308. | |
| PR | | 05-MAR-1999; | 99DK-0000309. | |
| PR | | 09-MAR-1999; | 99US-0123543. | |
| PR | | 10-MAR-1999; | 99US-0123623. | |
| PR | | 11-MAR-1999; | 99US-0123641. | |
| PR | | | 99US-0123642. | |

XX (NOVO) NOVO-NCRDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI
XX MPI: 2000-105851/09.
DR
XX N-PSDB; 245342.
XX
PT New mannases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions
XX
PS Claim 35; Page 224-225; 242pp; English.
XX
XX The present sequence represents a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation), in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hard or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.
CC
SQ Sequence 369 AA;

Query Match 41.2%; Score 1075; DB 21; Length 369;
Best Local Similarity 66.2%; Pred. No. 1.5e-72;
Matches 196; Conservative 37; Mismatches 63; Indels 0; Gaps 0

QY 34 SGEFYAGTLLYDANGNPFMRGINHGHAWYKDDATTAIEGIANTGVNRYVLSDDGQMT 93
DB 72 ngfyldgttllyangnqpfmrginbhsykhpleameadtgansitrvsddqqr 131
QY 94 KDDIHTVRNLISLAEENHLVAVPEVDATGYDSIASLRAVDYMIEMRSALICKEDTVII 153
DB 132 kdddevavakllsleakhsivaalevndaigtddleplktvdyvlelkdallgkckvll 191
QY 154 NIANEFEGWEGDADWDYKQAIPIPRRNAGLNTLWVDAAGKQFQSHDYGREVPNAD 213
DB 192 nlsnewfgswsgewadgykkaipllireaglkhtlnvdaagwqgfprslhekglevfnsd 251
QY 214 PQRNTFESIHMEYVGGNSOVRTNDRVLNODLAVIGEPGHRHNGVDDEATMISYSE 273
DB 252 plqnrtfslhmewaagngqgvkndagylekrlavrlgefighhygrdvaavtllshse 311
QY 274 QGQGVGLTANSMKNGNPEWETLDSNDWAANNLTAMGNTVNGPYGLRETSRLSTVF 329
DB 312 kydvglawslwngnsgvgyelldatctfstqgltegerlrvhngplketcelsvsvy 367

RESULT 9
Y54132
ID Y54132 standard; Protein; 320 AA.
XX
AC Y54132;
XX
DT 27-MAR-2000 (first entry)
XX
DE Amino acid sequence of a Bacillus sp. mannanase enzyme.
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;

| | | |
|----|---------------------------------------------------------------------------|---------------------------|
| XX | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | |
| KW | plant material degradation; recycled waste paper; paper making pulp; | |
| KW | guar; locust bean gum; mannan-containing food; coffee extract; | |
| KW | cleaning composition; machine washing; hard-surface cleaner; | |
| KW | dishwashing; oral; dental; contact lens; body-care composition; | |
| KW | fabric softener; oil well drilling; subterranean formation fracture. | |
| XX | | |
| OS | Bacillus sp. | |
| XX | | |
| EH | Key | Location/Qualifiers |
| FT | Peptide | 1..28 |
| FT | | /note- "signal peptide" |
| FT | Domain | 29..320 |
| FT | | /note- "catalytic domain" |
| XX | | |
| PN | W09964619-A2. | |
| XX | | |
| PD | 16-DEC-1999. | |
| XX | | |
| PE | 10-JUN-1999; | 99WO-DK00314. |
| XX | | |
| PR | 10-JUN-1998; | 98US-0111256. |
| PR | 20-OCT-1998; | 98DK-0001340. |
| PR | 20-OCT-1998; | 98DK-0001341. |
| PR | 28-OCT-1998; | 98US-0105970. |
| PR | 28-OCT-1998; | 98US-0106054. |
| PR | 23-DEC-1998; | 98DK-0001125. |
| PR | 05-MAR-1999; | 99DK-0000306. |
| PR | 05-MAR-1999; | 99DK-0000307. |
| PR | 05-MAR-1999; | 99DK-0000308. |
| PR | 05-MAR-1999; | 99DK-0000309. |
| PR | 09-MAR-1999; | 99US-0123543. |
| PR | 10-MAR-1999; | 99US-0123623. |
| PR | 10-MAR-1999; | 99US-0123641. |
| PR | 11-MAR-1999; | 99US-0123642. |
| XX | | |
| PA | (NOVO) NOVO-NORDISK AS. | |
| XX | | |
| PI | Kauppinen MS, Schnelein M, Schnorr K, | Andersen LN, Bjornvad ME; |
| XX | | |
| DR | WPI; 2000-105891/09. | |
| DR | N-PSDB; 245345. | |
| XX | | |
| PT | New mannanases for treatment of textiles, plant material and coffee | |
| PT | extract, and in cleaning compositions | |
| XX | | |
| PS | Claim 35; Page 228-229; 242pp; English. | |
| XX | | |
| CC | The present sequence represents a mannanase enzyme (also known as | |
| CC | mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). | |
| CC | The mannanase hydrolyses galactomannans. Specifically, mannanases | |
| CC | hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans, | |
| CC | glucomannans, and galactoglucomannans. The mannanase protein, or | |
| CC | preparations containing it, are used to improve properties of cellulostic | |
| CC | or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based | |
| CC | sizes or printing pastes). They are also used to degrade or modify | |
| CC | plant materials (particularly recycled waste paper, paper making pulps, | |
| CC | or material containing guar or locust bean gums (thickeners), or to | |
| CC | reduce viscosity of mannan-containing foods or feeds). The mannanases | |
| CC | are also used to process coffee extracts (to inhibit gel formation); in | |
| CC | cleaning compositions (for machine washing of fabrics, as hard-surface | |
| CC | cleaners, for hand or machine dishwashing, also in oral, dental, contact | |
| CC | lens or body-care compositions) where they remove mannan-containing | |
| CC | soils and prevent binding of some soils to cellulostics; and in fabric | |
| CC | softeners. They can also be used in oil well drilling to fracture | |
| CC | subterranean formations. | |
| XX | | |
| XX | Sequence 320 AA: | |

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 36.78; | Score 957.5; | DB 21; | Length 320; |
| Best Local Similarity | 58.48; | Pred. No. 7.1e-64; | | |
| Matches 181; Conservative | 42; | Mismatches 86; | Indels 1; | Gaps 1; |

| | | | |
|--------|---------------------------------------------------------------------------|----------------------------------------------------------------|-----|
| QY | 16 | LLASLIFVSGISTANANGSFYVSGTTLTYDANGNPFVWRGINHHGAMVKDQATTAISGIA | 75 |
| Db | 12 | lrvvvlvfyvsglaaagsgfhwkgfclldknngdpyvynghnswtkqgleeipala | 71 |
| QY | 76 | NTGANVFRVYLSGGGGWMTKDDIHVTNLTSLAEDNHLVAPEVHDATGDSIASINRAVD | 135 |
| Db | 72 | eegenvtrvrlsnvgqwekdaseelarvllaetelylttlevlnhatsgndpddakvxd | 131 |
| QY | 136 | YVIEIMRSALIGKEDVYIINANEMFEGSWGDAMDGKQALPRLNAGLNTLTAVDAAGW | 195 |
| Db | 132 | lytleamadvlykgtedrvyllnanevgyawrsdvwaaeqaprlrlsaglahtllyvdaegw | 191 |
| QY | 196 | GQFQFQSHDQGREVFVNADPQRNMSEIHHYEXYAGGNASQVRNIDRVYNODLALYISGEF | 255 |
| Db | 192 | gqypasfhergadvtasdpdknmfmslhmeyagdadavsenldgvaenlavvlgfeg | 251 |
| QY | 256 | HRHNGGVDDATIMSYSEQRCYVGLMWSKNGKPEWEYLDLSNDWAGNNLTAMGNTIYNG | 315 |
| Db | 252 | hrhdgdvddedallaylaeergvylawsvyngsgvreyldleqpsgpltswgerlvyg | 310 |
| QY | 316 | PYGLRSTSRLL325 | |
| Db | 311 | emgllkvldhl320 | |
| RESULT | 10. | | |
| ID | Y54130 | | |
| XX | Y54130 standard; Protein: 305 AA. | | |
| XX | Y54130; | | |
| DT | 27-MAR-2000 (first entry) | | |
| DE | Amino acid sequence of a partial <i>Bacillus</i> sp. mannanase enzyme. | | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | | |
| KW | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | | |
| KW | mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre; | | |
| KW | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | | |
| KW | plant material degradation; recycled waste paper; paper making pulp; | | |
| KW | guar; locust bean gum; mannan-containing food; coffee extract; | | |
| KW | cleaning composition; machine washing; hard-surface cleaner; | | |
| KW | dishwashing; oral; dental; contact lens; body-care composition; | | |
| KW | fabric softener; oil well drilling; subterranean formation fracture. | | |
| OS | <i>Bacillus</i> sp. | | |
| PN | WO9964619-A2. | | |
| PD | 16-DEC-1999. | | |
| PF | 10-JUN-1999; 99NO-DK00314. | | |
| PR | 10-JUN-1998; 98US-0111256. | | |
| PR | 20-OCT-1998; 98DK-0001340. | | |
| PR | 20-OCT-1998; 98DK-0001341. | | |
| PR | 28-OCT-1998; 98US-0105970. | | |
| PR | 28-OCT-1998; 98US-0106054. | | |
| PR | 23-DEC-1998; 98DK-0001725. | | |
| PR | 05-MAR-1999; 99DK-0000306. | | |
| PR | 05-MAR-1999; 99DK-0000307. | | |
| PR | 05-MAR-1999; 99DK-0000308. | | |
| PR | 05-MAR-1999; 99DK-0000309. | | |
| PR | 09-MAR-1999; 99US-0123543. | | |
| PR | 10-MAR-1999; 99US-0123623. | | |
| PR | 10-MAR-1999; 99US-0123641. | | |
| PR | 11-MAR-1999; 99US-0123642. | | |
| PA | (NOVO) NOVO-NORDISK AS. | | |
| PT | Kaupplen MS, Schnelein M, Schmitt R, Andersen LN, Bjornvad ME; | | |


```

Db      72 etgantvrlvlnbgpwekdasselarvlaetcyglttlevhdatgsdnpddldkavd 131
OY      136 YNEMRSALIGKEDFYINIANEMFEGSWEGDAMADGYKQALPRLRNAGINHTLWADA 192
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      132 yvlemadvkgtedrvllnlanewygaersdlwakayagaiprlrsaglahcllida 188

RESULT 12
P91007
ID      P91007 standard; protein: 487 AA.
XX
AC      P91007;
XX
DE      13-MAR-1990 (first entry)
XX
DE      Beta-mannase of alkalophilic Bacillus.
XX
KM      Escherichia coli.
XX
OS      Bacillus sp.
XX
PN      JP01228477-A.
XX
PD      12-SEP-1989.
XX
PF      09-MAR-1988; 88UP-0053774.
XX
PR      03-MAR-1988; 88UP-0053774.
XX
PA      (SHKJ) SHINGIJUTSU KAIHATSU.
XX
DR      WPI: 1989-312223/43.
XX
DR      N-PSDB: N91621.
XX
PT      Beta-mannase gene deoxyribonucleic acid
PT      - obt. by culturing transformed Escherichia coli
PT      contg. recombinant plasmid
PS
PS      Claim 1; Fig 3; 10pp; Japanese.
CC      E. coli transformants produce beta-mannase efficiently. It is obt.
CC      from Bacillus sp. FERM P-8856, -8857 and -8858.
XX
SQ      Sequence 487 AA:

Query Match      16.1%; Score 419; DB 10; Length 487;
Best local Similarity 27.0%; Pred. No. 1.6e-23;
Matches 142; Conservative 78; Mismatches 203; Indels 102; Gaps 22;

OY      25 SGTSTANANSGFYVSGTTLTDAN---GNPFVKNGINHGH---AMYKD---QATTAIE 72
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      4 ngsalnsnpnanqtknvswlanlpkshkrvvsghfsgysdclawlkqcarelsgkmp 63
OY      73 GIANTGANT--VRIVLSDGQWTKRDIHTVRNLSIAEDNHLVAV-----PEVHATGY 124
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      64 gliscdyknqgltllyadqisycn-----qellfnwngglvtlsvhmpdpfhsgey 118
OY      125 DSIASINRAVDY-----NIEKRSALIGKED-----TVIINIANEMFEGSWEDA 167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 ktllpsqfqlnqtrtegrtwkmdkmdagldelqngvctvlfpplhemgew---f 175
OY      168 W--ADGKQALPRLRNAGLN-----HTLMTDAAGGQFPPOSJHDYGREYFNADPO 215
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      176 wvgaeyngqdgtranaiaawrdmyqyftcherklmnlwvyspdyrdh---vtsyypg 232
OY      216 RNTMFESIHWMEYVAGNAGNSOVRTNIDRYLNODLALVIGEGHRHT--NGDVCEATIMSYSEQ 274
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      233 anyvdvalaisy--hpdphsiltqymmlajdkpfaiaeisppemagsfysyigaikq 291
OY      275 ---RGVGMWAMS--WK--GNGPEWEYLDLSNDWAGNNLTWAGNTIYNGPYGLBETSRLST 327
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      292 kyprtyvflawndkwspnhrgraw---dlnd-----sw 322

```

```

OY      328 VFTGGSGSGS--PTTYLDEGSGWGTGSSLS--GGFWATWENSSKSHSKADIQISS 384
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      323 vvnrgldyggspnatvlyldennlswsgeftggpwtensangtqslkadvlign 382
OY      385 NSOHLVHTQNTSLQONSRIQATVKHANMGSYGNGMTARLYKTKGTWVSGSFVPLNG 444
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      383 nsyh-1qktvnnlssfknlkshsswgnvsgymtarvfkfsgsawrmagelcgtag 441
OY      445 SSGTTLSDLSNWONTLQVREIGVQFQASDSSGQTSIYIDNVIV 489
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      442 krtalsldltkvnldhvtrelgveyekapansgkatalyldhvtv 486

RESULT 13
Y54131
ID      Y54131 standard; protein: 132 AA.
XX
AC      Y54131;
XX
DE      27-MAR-2000 (first entry)
XX
DE      Amino acid sequence of a partial Bacillus mannanase enzyme.
XX
KM      Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KM      endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KM      mannan; galactomannan; glucomannan; galactoglucomannan; cellulose; fibre;
KM      synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KM      plant material degradation; recycled waste paper; paper making pulp;
KM      guar; locust bean gum; mannan-containing food; coffee extract;
KM      cleaning composition; machine washing; hard-surface cleaner;
KM      dishwashing; oral; dental; contact lens; body-care composition;
KM      fabric softener; oil well drilling; subterranean formation fracture.
XX
OS      Bacillus clausii.
XX
XX      WO964619-A2.
XX
PD      16-DEC-1999.
XX
PF      10-JUN-1999; 99WO-DK00314.
XX
PR      10-JUN-1998; 98US-0111256.
PR      20-OCT-1998; 98DK-0001340.
PR      20-OCT-1998; 98DK-0001341.
PR      28-OCT-1998; 98US-0105970.
PR      28-OCT-1998; 98US-0106054.
PR      23-DEC-1998; 98DK-0001725.
PR      05-MAR-1999; 99DK-0000306.
PR      05-MAR-1999; 99DK-0000307.
PR      05-MAR-1999; 99DK-0000308.
PR      05-MAR-1999; 99DK-0000309.
PR      09-MAR-1999; 99US-0123543.
PR      10-MAR-1999; 99US-0123623.
PR      10-MAR-1999; 99US-0123641.
PR      11-MAR-1999; 99US-0123642.
XX
PA      (NOVO) NOVO-NORDISK AS.
XX
PI      Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX      WPI: 2000-105891/09.
XX      N-PSDB: Z45344;
XX
PT      New mannanases for treatment of textiles, plant material and coffee
PT      extract, and in cleaning compositions
XX
PS      Disclosure; Page 227; 242pp; English.
XX
CC      The present sequence represents a mannanase enzyme (also known as
CC      mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC      The mannanase hydrolyses galactomannans. Specifically, mannanases
CC      hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,

```


XX WPI: 1997-402610/37.
DR N-PSDB: T77055.

PT Desizing cellulose⁶-containing fabric or textile using an enzyme
hybrid - which comprises a catalytically active amino acid sequence
PT of a non-cellulolytic enzyme linked to an amino acid sequence
PT comprising a cellulose binding domain.

XX Example 2; Page 52-54; 72pp; English.

CC This protein comprises the alkaline cellulase (endoglucanase) of
CC Bacillus agaradherens NCIMB 40482. It was expressed in Bacillus
CC subtilis PL2306 transformants following PCR amplification (see
CC T77053-54) of B. agaradherens genomic DNA and ligation of the PCR
CC product into vector pDN1981. DNA encoding the cellulose binding
CC domain (CBD) of the alkaline cellulase was subsequently amplified
CC (see T77056-57) for use in the construction of a novel alpha-
CC amylase-CBD hybrid enzyme (see T77058). A claimed process for
CC desizing cellulose-containing fabric or textile comprises treating
CC the fabric or textile with a modified enzyme (enzyme hybrid)
CC comprising a non-cellulolytic enzyme linked to a CBD. The process
CC gives improved enzyme performance by modifying the enzyme so as to
CC increase its affinity for cellulosic fabric.

SQ Sequence 400 AA;

Query Match 8.8%; Score 229.5; DB 18; Length 400;
Best Local Similarity 23.4%; Pred. No. 1.5e-09;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

QY 6 KRFEFTSLSLASSSTLEFSGSTANANS-----GPFVSGTLYDANGNPFVNRGI-NH 58
DB 2 Kktltfvtvlmvtvalfsygnltacndsvveehqqlsngelvnergeqvqlkqmsn 61
QY 59 GRAMY-----KDQATTAIEGANTGANTVRIVL--SDGGQWTKDI-HTVRNLT 103
DB 62 glqwygqfvnyesmkvlrd-----dwglvfrfaamytsagylddpsvkekkea 111
QY 104 ISIAEDNHLVAVPEVDATGYDSIASLNRAVDYWIEMKRSALIGKEDTVITINANEFGSW 163
DB 112 veealldldlyldwhlisdndpnlykeekdfidem-selygdypnvilyelanepgs- 169
QY 164 EGDWADGVRK---QAIPRLRNAGLWHTLWVDAAGQGQFPOSIHGGRVFNADPQRNTW 219
DB 170 -dvtwgnqlkpyaeewipilrindpnlliv---gctwsqdvh-haadnqladp--nym 222
QY 220 FSIHMYEYAGNASQVRTNIDRYLNODLALVIGFGRHTNGD---VDEATI-MSYSEQ 274
DB 223 yafhf--yagthgqnrlrdqvdalqgaalfvewgtsaatggyvfldeagwldtmde 280
QY 275 RGYGMLAWS-----WKGNQPEWEYLDLSNDMAGNNLTMGNTIVNGPYGLRETS 323
DB 281 rnlswanwsltlhdeessaalmpgarp-----tgyvteaelspsgltfvre--klres 330
QY 324 RLSTVFTGGSDGSGTPTTLYDFEGSMQGTGSSL-----SGGPWATVEMS 369
DB 331 slppsdptppsdgppdpptppsdggyeyawdpdqilyrnelvymgqlwgakwlt 384

Search completed: December 19, 2000, 16:49:31
Job time: 1181 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:41:28 ; Search time 156.76 Seconds
(Without alignments)
52.392 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LUNGFKIRISITLSLILASS.....QASDSSGQTSTIINDVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 229.5 | 8.8 | 400 | 2 | US-08-713-298B-2 |
| 2 | 229.5 | 8.8 | 400 | 2 | US-08-870-180B-2 |
| 3 | 229.5 | 8.8 | 400 | 3 | US-08-814-052-4 |
| 4 | 229.5 | 8.8 | 400 | 3 | US-08-812-829-4 |
| 5 | 229.5 | 8.8 | 462 | 2 | US-08-870-180B-13 |
| 6 | 213 | 8.2 | 467 | 2 | US-08-727-548-2 |
| 7 | 189 | 7.2 | 551 | 2 | US-09-033-537A-1 |
| 8 | 174 | 6.1 | 430 | 2 | US-08-924-440-2 |
| 9 | 160 | 6.1 | 531 | 2 | US-07-862-588B-7 |
| 10 | 143 | 5.5 | 521 | 1 | US-08-276-213-3 |
| 11 | 142 | 5.4 | 358 | 1 | US-08-604-913B-11 |
| 12 | 129 | 4.9 | 1536 | 1 | US-08-038-662-2 |
| 13 | 129 | 4.9 | 1536 | 1 | US-08-302-832-2 |
| 14 | 129 | 4.9 | 1536 | 2 | US-08-530-198-2 |
| 15 | 129 | 4.9 | 1536 | 2 | US-08-469-880-2 |
| 16 | 129 | 4.9 | 1536 | 2 | US-08-728-470-2 |
| 17 | 129 | 4.9 | 1536 | 2 | US-08-617-697-2 |
| 18 | 123.5 | 4.7 | 880 | 3 | US-08-947-965-77 |
| 19 | 116.5 | 4.5 | 983 | 3 | US-08-164-292B-26 |
| 20 | 116.5 | 4.5 | 983 | 3 | US-08-845-623-26 |
| 21 | 116.5 | 4.5 | 983 | 3 | US-08-815-927-26 |
| 22 | 112.5 | 4.3 | 490 | 1 | US-08-672-571A-1 |
| 23 | 111 | 4.3 | 574 | 3 | US-08-732-433-1 |
| 24 | 109 | 4.2 | 735 | 2 | US-08-313-185-48 |
| 25 | 109 | 4.2 | 735 | 2 | US-08-459-499-9 |
| 26 | 108 | 4.1 | 1545 | 4 | PCT-US95-10661A-4 |
| 27 | 107.5 | 4.1 | 666 | 2 | US-08-737-716-14 |
| 28 | 107.5 | 4.1 | 1222 | 2 | US-08-682-517-15 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 29 | 107.5 | 4.1 | 1252 | 2 | US-08-682-517-9 | Sequence 9, Appl |
| 30 | 106.5 | 4.1 | 750 | 5 | 5457037-3 | Patent No. 5457037 |
| 31 | 106 | 4.1 | 1338 | 2 | US-08-728-470-9 | Sequence 9, Appl |
| 32 | 106 | 4.1 | 1599 | 2 | US-08-617-697-9 | Sequence 9, Appl |
| 33 | 105.5 | 4.0 | 751 | 5 | 5457037-5 | Patent No. 5457037 |
| 34 | 105 | 4.0 | 385 | 2 | US-08-387-942C-26 | Sequence 26, Appl |
| 35 | 105 | 4.0 | 654 | 1 | US-08-392-828C-2 | Sequence 2, Appl |
| 36 | 105 | 4.0 | 654 | 3 | US-09-330-945-2 | Sequence 4, Appl |
| 37 | 105 | 4.0 | 997 | 2 | US-08-387-942C-4 | Sequence 4, Appl |
| 38 | 104 | 4.0 | 398 | 2 | US-08-853-659A-45 | Sequence 12, Appl |
| 39 | 104 | 4.0 | 735 | 2 | US-08-459-499-12 | Sequence 12, Appl |
| 40 | 103.5 | 4.0 | 553 | 2 | US-08-661-052-16 | Sequence 15, Appl |
| 41 | 103 | 4.0 | 434 | 2 | US-08-989-925-1 | Sequence 1, Appl |
| 42 | 103 | 4.0 | 1702 | 4 | PCT-US95-10661A-5 | Sequence 5, Appl |
| 43 | 102.5 | 3.9 | 623 | 3 | US-09-041-991A-6 | Sequence 6, Appl |
| 44 | 101.5 | 3.9 | 566 | 2 | US-07-862-588B-4 | Sequence 4, Appl |
| 45 | 101.5 | 3.9 | 928 | 1 | US-08-474-140-11 | Sequence 11, Appl |

ALIGNMENTS

RESULT 1
US-08-713-298B-2
; Sequence 2, Application US/08713298B
; Patent No. 5922586
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dammann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjoernvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; TITLE OF INVENTION: Cellulolytic Enzymes
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59225860 No. 5922586disk of No. 5922586th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,298B
; FILING DATE: 13-SEPT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 3794,424-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 867-0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-713-298B-2

Query Match 8.8%; Score 229.5; DB 2; Length 400;
Best Local Similarity 23.4%; Pred. No. 5.0e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

US-08-870-180B-2
; Sequence 2, Application US/08870180B
; Patent No. 5945327
; GENERAL INFORMATION:
; APPLICANT: Oultup, Helle
; APPLICANT: Dammann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg, Ed-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; TITLE OF INVENTION: Cellulytic Enzymes
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5945327disk of No. 5945327th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,180B
; FILING DATE: 6-June-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 3794.434-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-870-180B-2

Query Match 8.8%; Score 229.5; DB 2; Length 400;

Best Local Similarity 23.4%; Pred. No. 5,8e-12;

Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

US-08-870-180B-2
; Sequence 4, Application US/08814052
; Patent No. 6015783
; GENERAL INFORMATION:
; APPLICANT: von der Osten, Claus
; APPLICANT: Cherry, Joel R.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Wind, Jesper
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
; TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 60157830 No. 6015783disk of No. 6015783th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814,052
; FILING DATE: 06-MAR-1997
; CLASSIFICATION: 510
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambitis, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4684.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:

```

RESULT 4
US-08-812-829-4
Sequence 4, Application US/08812829
Patent No. 6017751
GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Bjornvad, Mads E.
APPLICANT: Wind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60177510 No. 6017751disk of No. 6017751th America, Inc
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,829
FILING DATE: 06-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambitis, Elias J

RESULT 5
 US-08-870-180B-13
 Sequence 13, Application US/08870180B
 Patent No. 5945327.
 GENERAL INFORMATION:
 APPLICANT: Outtrup, Helle
 APPLICANT: Damdbmann, Claus
 APPLICANT: Olsen, Arne
 APPLICANT: Bisq rd-Frantzen, Henrik
 APPLICANT: Sch lein, Martin
 APPLICANT: J rgensen, Per
 APPLICANT: Bjørnervad, Mads
 TITLE OF INVENTION: DNA Constructs and Methods of Producing
 TITLE OF INVENTION: Cellulytic Enzymes
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 59453270 No. 5945327/disk of No. 5945327th America, Inc
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk


```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,180B
FILING DATE: 6-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794,434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-180B-13

```

```

Query Match      8.8%; Score 229.5; DB 2; Length 462;
Best Local Similarity 23.4%; Pred. No. 7.3e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

```

QY 6 KRTFTSLSLASSILFVSGTSTANANS-----GFTVSGTLYDANGNPFVARGI-NH 58
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KRTFTTIFVLMVLFISGNTTAADNDVVEEHQSLISNGELVNEGEQVOLKGMSSH 61
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 GHAWY-----KQDATAIEGIANTGANTVRIYV--SDGGQWTKDDI-HYVRNL 103
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 GLQMTGCFVYVESKMKLKD-----DWGIVFPAAMTSSGGYIDDPVYKEKYEKA 111
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 ISLAEDNHLVAPEVHDATGYDSIASLNRAVDYIEMRSALIGKEDTVIINIANEMFGSW 163
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 112 VEAALDIDVILDMHLSNDNPRIYKEAKDFEDM--SELYGDYPRVIVIEINERPGS- 169
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 EGAMADGIR-----QAIPRLRNAGLNHTLVDAAGKGFQPSIHIDYGREYFNADPQANTM 219
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 -DVTWGQIPIYAEVYPIIRNNDPNNIIV--GTGTWSQDVH--HAADNQLADP--NVN 222
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 FSIHMYEAGNNAOVRTNIDRVNLQDLALVIGFGRHTNGD---YDEATI--MSYSEQ 274
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 223 YAHF--YAGTHGQNLRDQVDYALDGAALFVSEMGTSATGSGVFLDERQVWIDPFMD 280
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 RGVGWLAMS-----WKGNGPEWEYLDLSNDMAGNNLTAMGNTIVNGPYGLRETS 323
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 RNLNMANWMSLTHKDESSAALMPGANP-----TGWTAEALSPSGTFVRE--KIRESA 330
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 RLSTVTFGGSGDGTSTTIYDEGSMQGTGSSL-----SGGWAVENTMS 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 331 STPPSDPTPPSDGEPDPTPPSDPGKYPADPQOITYNEIYVHNGQLMOAKMWT 384
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
US-08-727-548-2
; Sequence 2, Application US/08727548
; Patent No. 5856165
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,548
; FILING DATE:

```

```

CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-727-548-2

```

```

Query Match      8.2%; Score 213; DB 2; Length 467;
Best Local Similarity 22.8%; Pred. No. 2e-10;
Matches 99; Conservative 62; Mismatches 157; Indels 116; Gaps 22;

```

```

QY 6 KRTFTSLSLASSILFVSGTSTA-----NANGFTVSGTLYDANGNPFVARGI-N 57
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KRTFTTIFVLMVLFISGNTTAADNDVVEEHQSLISNGELVNEGEQVOLKGMSS 60
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 GHAWYKQDATAIEGIANTGANTVRIYVSDGGQWTKDD-----IHT----- 99
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 HGLQMTG-----OFVYVESKMKLRDMDGITVFPAAMTSSGGYIDDP 102
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 100 -----VRNLISLAEDNHLVAPEVHDATGYDSIASLNRAVDYIEMRSALIGKEDTVIIN 154
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 SVYKEKVEVEAALDGIYVILDMHLSNDNPRIYKEAKDFEDM--SELYGDYPRVIVIE 161
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 IANEMFGS--WEGD--ANADGKQALPRLRNAGLNHTLVDAAGKGFQPSIHIDYGREY 209
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 IANEPGSGSVTMDNQIKPFAE--EVIPIVRNDPNNIIV--GTGTWSQDVH--HAADN 214
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 FNADPQNTMFSIHMYEAGNNAOVRTNIDRVNLQDLALVIGFGRHTNGD---YDE 265
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 215 QLADP--NMYIAHF--YAGTHGQNLRDQVDYALDGAALFVSEMGTSATGSGVFLDE 270
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 AII-MSYSEQRGVWLAWS-----WKGNGPEWEYLDLSNDMAGNNLTAMGNTIV 313
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 AQWVIDFMDERNLNMANWMSLTHKDESSAALMPGANP-----TGWTAEALSPSGTFV 322
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 NGPYGLRETSKLSYV-----TGGSPDGTSTTIYDEGSMQGTGSSL----- 358
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 323 -----REKIRRSASITPPSDPTPPSDGEPDPTPPSD--PGKYPADPMSNQITYNEI 375
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 ---SGGWAVENTMS 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 376 VYHNGQLMOAKMWT 389
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Shlehn, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A

```

FILING DATE: 02-MAR-1998
 CLASSIFICATION: 008
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 0993/95
 FILING DATE: 08-SEP-1995
 APPLICATION NUMBER: PCT/DK96/00364
 FILING DATE: 03-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Green, Reza
 REGISTRATION NUMBER: 38,475
 REFERENCE/DOCKET NUMBER: 4492.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 551 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-033-537A-1

Query Match 7.2%; Score 189; DB 2; Length 551;
 Best Local Similarity 21.3%; Pred. No. 2.9e-08;
 Matches 108; Conservative 72; Mismatches 181; Indels 146; Gaps 24;

QY 38 VSGTLLYDANGNPFVARGI-NHGNAVYKQOATTAIGIANTGANTVRIYLSDGQWTKDD 96
 DB 11 VQGNQVGGSGQAVOLVGGSHGLQWYGN-----FYKSKSLQMRNN 52
 QY 97 --IHVYRNILSLAEDNHLVAVEVDATGIDYSLASLRA-----VDYWI----- 138
 DB 53 WCINFRAMAYRAEDGY-ITDPSVKKKVKEAVQASIDLGIVYIDWHIISDGNPTYKAQ 111
 QY 139 -----EMRSALIGKEDTYIINIANEMFGSMGDAAD--GY-KQALPRLNAGLNTLWY 190
 DB 112 SKAFQEMATLGNPNVYIEIANEPNGN--VSMADVSTAEVYTAIRADPDGVVYV 168
 QY 191 DAGMGQFPOSIDHYGREYFNADPORNTMFSIHMEYAGNASSOVRTNIDRYLNDLALV 250
 DB 169 GSPPTMSQ-----DIHLAADNPVSHSNVYKALHF--YSGTHGFLLDRITTYANKGAALF 220
 QY 251 IEEFGHRTNGD-----VDEATIMSYSEQRGVGWLAMS-----WKGNGPEWEYL 294
 DB 221 VTEMGSDASGNGGPFYPOSKIEWIDFLNARKISVMNMSLADKVEYSAALMFGASP----- 275
 QY 295 DLSNMGANNLWANGNTIYNGPYGLRETSRLSTVFTGGGSDGT---SPTTLYDFEGSMQ 351
 DB 276 --TGGWTDALQLESQKMW-----RDQIRQA--TGGSGGNFTAPAAPNLNLSATAGNAQ 323
 QY 352 -GMTGSSLSGGFWAVTEMESGSHSLKADQLQSSNQHYLYQNTSLQONSHRIQATVYH 410
 DB 324 VSLTNNAVSGAR-----SYTVK-----RATISG 346
 QY 411 ANMGSYGNCMTARLYYKIG--HGTYWY-----SGSFVPINGSSGTTLSLDSNVQNLSCV 463
 DB 347 GPYTNVATGVTATSYNTGLTNGTLYYVVASNS-----AGSSANSAQAASAPASGAST 402
 QY 464 REIGVOFO-----SASDSSGQTSIYIDN 486
 DB 403 GNLVVOYKVGDTSATDNQMKPSFNKN 429

RESULT 8
 US-08-924-440-2
 Sequence 2, Application US/08924440
 Patent No. 5871550
 GENERAL INFORMATION:
 APPLICANT: Frits et al.
 TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
 NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International, Inc.
 STREET: 925 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/924,440
 FILING DATE: August 27, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Christopher L. Stone
 REGISTRATION NUMBER: 35,696
 REFERENCE/DOCKET NUMBER: GC388
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 845-6504
 TELEFAX: (650) 845-7555
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 430 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-924-440-2

Query Match 6.7%; Score 174; DB 2; Length 430;
 Best Local Similarity 24.7%; Pred. No. 4e-07;
 Matches 81; Conservative 49; Mismatches 130; Indels 68; Gaps 19;

QY 26 GSTANANSQ-FYVSGTLLYDANGNPFVARGIN-HGHNAVYKQOAT-TALEGIA-NTGANT 81
 DB 124 GGTGTPERYGKQVQVQICLDEHGNPVQRMSTHGIQWFDHCLTDSIALAYDKADI 183
 QY 82 VRL---VLSDDGQW---KDIHYRNILSLAEDNHLVAVEVDATGIDYSLASLRAVY 134
 DB 184 IRLSMYIOEDGYETINRGFTDRH---QLIDVATNGLVYIVDWHILTFGDPDHYNDLRAK 240
 QY 135 DWIEMRSALIGKEDTYIINIANEMFGSMGDAAD--GY-KQALPRLNAGLNTLWY 191
 DB 241 TFFAEIAQNHASKTN-VLEIANE-----PNGVSMASIKSYAEVLPVIRQRPDSYITVG 295
 QY 192 AAGMGQFPOSIDHYGREYFNADP--QRNTMFSIHMEYAGNASSOVRTNIDRYLNDLALV 249
 DB 296 TGMWSLGVY-SGSGPAEIAAPVNASNIMYAFHFY-----AASHRDYVNLNLRASL 348
 QY 250 ---VIGFGRHRTNGD-----VDEATIMSYSEQRGVGWLMSK----- 285
 DB 349 FPFVYTFEFTETTYTGANDFQMAADR--YIDMAERKIGMTKWNYSDDPRSGAVTQPGTC 406
 QY 286 -GNGPEWEYLDLSDWAGNNLWANGNTI 312
 DB 407 ASGGP-----WSSSLKASGQWV 424

RESULT 9
 US-07-862-588B-7
 Sequence 7, Application US/07862588B
 Patent No. 5916796
 GENERAL INFORMATION:
 APPLICANT: Joergensen, Per Linaa
 APPLICANT: Sch. Iein, Martin
 APPLICANT: Hansen, Christian
 TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
 NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796d1sk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B

FILING DATE: 19920727
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: DR 164/90

FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.

REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids

TYPE: AMINO ACID
TOPOLOGY: linear

MOLECULE TYPE: protein
US-07-862-588B-7

Query Match
Best Local Similarity 6.1%; Score 160; DB 2; Length 531;
Matches 109; Conservative 73; Mismatches 214; Indels 132; Gaps 24;

11 ITLSLLASSLLEFVSGISTANANGS-----FVSGTITIDANGNPFVARGI-NHGHAMK 64
12 LVLEFVWLMGLLPVAGKGYAAPVPGOLKVGQNLVGGSGAVQVQVSSSHGLQMTG 71
65 DOATAIGIANTGANTVRIYLSGGQWTKDD--HTVRNLISLAEDNHLVAPEVHDAT 122
72 N-----FVKKSSLOQMRDNGINVRAMATSEDDY-ITDPSVKNKV 112
123 GYDSIASLNR-----VDYWI-----EMRSALIGKEDTYIINIANEWF 161
113 KEAVQASIDIALYVLIIDWHILSDGNPTKYAKSKAFQEMATLYGNTPNVIYEATSPTE 172
162 SWEQAMADGYQALPRLRMAGLNTLWVDAWGQFQSHDYREVFNDPQRTMFS 221
173 CVLGRQSS--BEVITATISIDPDGVVYVGSPTWSQ-----DILADNPVSHNVAYA 224
222 IHMYEYAGNASQVNTIDRVLNODLAVIGEGFGRHNTGNDVDEATIMSYSEQRGVWLA 281
225 LHF--YSGTHGQFLDRITRYANKGALIPVTEMGTSASGN-----263
282 WSMKNGPMEYELDSNDY---AGNNLTAMGNTIYNGPIGLAETSRLSVFTGGSSDGT 338
264 -----GGP-----YLPOSKKEIDFLNARKISWV-----WSLADKVETSAALMPGASPTGA 309
339 SPTTILDFEGSMQGTGSSLSGGPMNAVTEMSSKGSHTL-KADIQSSNSOHTLHVIQNTS 397
310 GP-----MPCNRMGKSGSSSN-----PASMWRQGNPTAALPYNLSANGN-----AQVS 355
398 LQONSRIQA--TVKHAN-----WGSVAGNMTARLYVKTG--HGYTWY-----SSSEFPI 442
356 LFMNNAVSAGATSVYKRAATISGQPYTNVDRGVATATSTYNTGLNGTLYYVVRASNS--- 411

443 NGSSGTTLSLDSLVNVLQVREIGVQFQ-----SADSSGQTSITIDN 486
412 AGSSANSAQASATPASGASATGNLVVQYKVDISATDNQKPSFNKN 459

RESULT 10
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655.

GENERAL INFORMATION:
APPLICANT: Thomas, Steven

APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael

TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard
CITY: Golden

STATE: CO
COUNTRY: USA

ZIP: 80401-3393
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213

FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna

REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08

TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000

TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids

TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHEICAL: NO
ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match
Best Local Similarity 5.5%; Score 143; DB 1; Length 521;
Matches 124; Conservative 70; Mismatches 180; Indels 216; Gaps 36;

32 ANSGF-VVSGTITIDANGNPFVARGIN-----HGHAHYKQDATTAEIGANTGA 79
1 AGGCIWHTSGRELLDANNVVRAGINWGEFENCNTVYG-LMSRDY-RSMLDQIKSLAY 58
80 NVRIYVSDGQWTKDDI--HTVRNLISLAEDNHLVAPEVDAAGYDSIASLNRNAVY 136
59 NTRLPYS-----DDLKPGTTPNSINFYQNN-----ODLOGTLISQYMDIVAVI 103
137 -----WIEMRSALI-----GKEDTYIINIANEW 159
104 AGQIGLRIILDRHRPDCSGSALMTYSSVEATWISDLOALQARYGNPTVVGFDLHNE 162
160 FGSWEQDAMADG-----YKQALPRLRNA--GLNHTLAV-----DAAGH----- 195
163 --PHDPACWGGGSDSIDMRLAERAGNAVLSVNPMLLIIFEGVQSYNGDSYWMGCLQGA 220
196 GQFP-----QSHIDYGREVFNDPQRTMFSIHMYEYAGNASQV-RINIDRVLN 244

Wed Dec 20 15:35:21 2000

us-09-339-159-2.ra1

Page 8

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match 4.9%; Score 129; DB 1; Length 1536;
Best Local Similarity 21.1%; Pred. No. 0.021; Indels 136; Gaps 21;
Matches 115; Conservative 74; Mismatches 221;

1 LNNFKKIFSTLTLASSILFVSGTSTANANGS-----FYVSGTLLYDANGNPFV 52
DB 881 INSG-----NLTAGNIVNINAGNLVESNANFKATNFTFNVGG--LFDNKGNSNI 929
QY 53 MRGINHGAWYKDOATTAEGIANGTANTVRIYLSGGGOWTODIHTVRNLISLAEDNH 112
DB 930 --SIAGGARFKIDNSKNLSTTNSSSTYRTIISGNITKNKNDL---NITNKGSDTEM 983
QY 113 VAVPEVHATGYDSIAS--LNRAVDYWIEMRSALIGKEDTVIINIANEWFSGWEGDAMAD 170
DB 984 QIGGDVSOKEGNEFTISSDKIN-----ITKQITIKAGVDGE-----NSDSD 1023
QY 171 GYKQAIPLRNAGLNLHTLVDAWGQFPOSIDYGEVFNADPQRTMFSIHMYEYAG 230
DB 1024 ATNNANLTIKTKELKLDLNLISGFNK-----AETAKDSDLTIGTNTNSAD--GT 1072
QY 231 NASQVTRINDRYLNDLALVIGFGRHRTNGVDVDEATIMSSEORGWGLAMSKNGNPE 290
DB 1073 NAKKV-----TFNQVKDSKISADGK-----VTLHSEKVTSG-----SNNMTE 1110
QY 291 WEYLDLSDMAG-----NLTAMGNTIVNGPYGLRETSRLSTVFTG----- 331
DB 1111 ----DSSDNNAGLTIDAKNTVNNNITSHKAVSISATSGEITTKTGTINATTGNVEITA 1166
QY 332 -----GSGDGTSPITLYDEFGSMOGWTSLSGGPVAWTESM-----KGSHS 375
DB 1167 QTGSIIGIESSGSVTLTATEGAL--AVSNISGNTVYTVANGALTTLAGSTIKGTES 1223
QY 376 LKADIOLS-----SNSQHYLVHIONTSLOONSRIOATVKHAN---WGSVNGMTARLY 425
DB 1224 VTTSSQSDIGTISGCTVEKATESLTQSNKIKATGCEANTVSATGIGTISGNTV 1283
QY 426 VKTGHGYTWYSGFVPINGSSG---TTLSDLSNVONLSQVREIGVQFOSASDSSGOTS 481
DB 1284 NVTANAGDLTVNGAEINATEGAATLTSSGKLTTEASHTSAKGVNLSADDSVAGS 1343
QY 482 IYIDNV 487
DB 1344 INAAV 1349

RESULT 13
US-08-302-832-2
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berleissner, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536; amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 4.9%; Score 129; DB 1; Length 1536;
Best Local Similarity 21.1%; Pred. No. 0.021; Indels 136; Gaps 21;
Matches 115; Conservative 74; Mismatches 221;

1 LNNFKKIFSTLTLASSILFVSGTSTANANGS-----FYVSGTLLYDANGNPFV 52
DB 881 INSG-----NLTAGNIVNINAGNLVESNANFKATNFTFNVGG--LFDNKGNSNI 929
QY 53 MRGINHGAWYKDOATTAEGIANGTANTVRIYLSGGGOWTODIHTVRNLISLAEDNH 112
DB 930 --SIAGGARFKIDNSKNLSTTNSSSTYRTIISGNITKNKNDL---NITNKGSDTEM 983
QY 113 VAVPEVHATGYDSIAS--LNRAVDYWIEMRSALIGKEDTVIINIANEWFSGWEGDAMAD 170
DB 984 QIGGDVSOKEGNEFTISSDKIN-----ITKQITIKAGVDGE-----NSDSD 1023
QY 171 GYKQAIPLRNAGLNLHTLVDAWGQFPOSIDYGEVFNADPQRTMFSIHMYEYAG 230
DB 1024 ATNNANLTIKTKELKLDLNLISGFNK-----AETAKDSDLTIGTNTNSAD--GT 1072
QY 231 NASQVTRINDRYLNDLALVIGFGRHRTNGVDVDEATIMSSEORGWGLAMSKNGNPE 290
DB 1073 NAKKV-----TFNQVKDSKISADGK-----VTLHSEKVTSG-----SNNMTE 1110
QY 291 WEYLDLSDMAG-----NLTAMGNTIVNGPYGLRETSRLSTVFTG----- 331
DB 1111 ----DSSDNNAGLTIDAKNTVNNNITSHKAVSISATSGEITTKTGTINATTGNVEITA 1166
QY 332 -----GSGDGTSPITLYDEFGSMOGWTSLSGGPVAWTESM-----KGSHS 375
DB 1167 QTGSIIGIESSGSVTLTATEGAL--AVSNISGNTVYTVANGALTTLAGSTIKGTES 1223
QY 376 LKADIOLS-----SNSQHYLVHIONTSLOONSRIOATVKHAN---WGSVNGMTARLY 425
DB 1224 VTTSSQSDIGTISGCTVEKATESLTQSNKIKATGCEANTVSATGIGTISGNTV 1283
QY 426 VKTGHGYTWYSGFVPINGSSG---TTLSDLSNVONLSQVREIGVQFOSASDSSGOTS 481
DB 1284 NVTANAGDLTVNGAEINATEGAATLTSSGKLTTEASHTSAKGVNLSADDSVAGS 1343
QY 482 IYIDNV 487
DB 1344 INAAV 1349

RESULT 14

US-08-530-198-2
Sequence 2, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME II, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530.198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

Query Match 4.9%, Score 129, DB 2, Length 1536;
Best Local Similarity 21.1%, Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;

1 LUNGFKKIFSLILSLASSILFVSGTSTANANG-----FYVSGTLLYDANGNPFV 52
881 INSG-----NLTAGNININAGNLTVESNAFKATINTFTVNG--LFDKNGNSNI 929
53 MRGINHGAWYKQDATTAEIGIANTGANTVRIVLSDGGQWTDIHTVNLISLAEDNL 112
930 --SIKAGARFDIDNSKNLSITNSSSYRTISGNITKNKGD---NINNESDTEFM 983
113 VAVPEYHATGDSIAS--LNPAYDIWEMRSALIGKEDVTIINIANEWTGSGWGDAMAD 170
984 QIGGVSOKEGELTITSSDKIN-----ITKQITIRAGVGE-----NSDSD 1023
171 GYKQAIPLRNALNHTLPLVDAAGQFPQSIHDYGREVFNDPQNTWFSIHMYEYVGG 230
1024 ATNNANLTIKTEKLITODLINSFKN-----AETIADGSDLTIGNTNSAD--GT 1072
231 NASQVRTNIDRYLNDOLALVYGEFGRHRTNGDYDEATINSYSEQRGVGLAWSKNGNPE 290
1073 NAKVY-----TFNQVSKRISADGKH-----VTLHSKVTESG-----SNNMTE 1110
231 WEYLDLSNDMAG-----KNLTAMGNTIVNGPYGLRETSRLSTVFTG----- 331
1111 ---DSDNNALTLTDARKNVYNNNITSKRAVISATSEITTKGTITINATGVEIRA 1166
332 -----GSGDGTSPITLYDFEGSGMQGWTGSSLSGGPNAVTEMS-----KGSMS 375

DB 1167 QNCSILGIESSGSSVLTALTEGAL---AVSNISGNTVTVANSQALTLTLAGSTIKGTES 1223
QY 376 LKADIQLS-----SNSQHYLVIONTSLOQNSRIQATVKAN---NGSVNGKATRLY 425
DB 1224 VTTSSQSGDIGGTIGGIVYEVATSESLTQNSNKKIKATTEGANTVSANGTIGTISGTV 1283
QY 426 VATGIGTYMGSFYPINSSG-----TTLSDLSNVQNLQSVRELVGFQFQSSDSSGQT 481
DB 1284 NYTANAGDLYVNGAEINATEGATLTYSGRKLTTEASHITSANGOVNLSHQDSVAGS 1343
QY 482 IYIDNV 487.
DB 1344 INANV 1349.

RESULT 15
US-08-469-880-2
Sequence 2, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469.880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bernstreser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:Vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 4.9%, Score 129, DB 2, Length 1536;
Best Local Similarity 21.1%, Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;

1 LUNGFKKIFSLILSLASSILFVSGTSTANANG-----FYVSGTLLYDANGNPFV 52

Db 881 INSG-----NLTAGNTVNTAGNTLVESNANFEAITNFTENVGG--LFDNKGNSNI 929
QY 53 MRGINHGHWYRDQATTAIEGIANTGANTVRIVSDGQWTRDIDHYRNIISLAEDNHL 112
Db 930 --SIKKGARFRDINDSKNLSTTNSSSTYRIIISGNTNKGDL---NITNGSDTEM 983
QY 113 VAVPEVHADTGDSIAS--LNRADVWTEMRBALIGKEDTVIINANEMFGSWEGDAMAD 170
Db 984 QIGGDVSOKEGNLITSSDKIN-----ITKQITINAGVDG-----NSDSD 1023
QY 171 GYKOAIPRLRNAGLNTHTLVDAAGWGPQS1HDYGREYFNADPQANTMESTIMTEYAGG 230
Db 1024 AITNANLITIKETELTODLNISGFNK-----AETAKGSDLTIGNTNSAD--GT 1072
QY 231 NSQVPTNIDRYLNODLAVIGFGRHRTNGVDENTIMYSQORGWGLAMSKNGNPE 290
Db 1073 NAKKV-----TFNQYKDSKISADGK-----VTLHSHKVEYTSG-----SNNTE 1110
QY 291 WEYLDLSNDWAG-----NNLTAWGNTIVNGPYGLRSTRSLSTVFTG----- 331
Db 1111 ---DSSDNAGLITIAKANTVYNNITSHKAVISATSGEITTKGTITINATTGNVEITA 1166
QY 332 -----GSGDGTSPPTLYDFEGSMQWGTSSISGCPNAVTEWS-----KGSHS 375
Db 1167 QTSILGIESSSGSVTLTATEGAL--AVSNISGNTVTVTANSALTTLAGSTIKGTES 1223
QY 376 LKADIQLS-----SNSQHLHYIOMTSLOONRRIQATVKHAN---WGSVGNMTARLY 425
Db 1224 VTTSSQSGDIGTISGGEVEVKATEBSLTTOSNKRKATIGEANVTSATGTIGGTISGNTV 1283
QY 426 VKTGHGYTWYSGSFVPINGSSG---TTLSDLSNVQNLQVREIGVQFQASDSSGQTS 481
Db 1284 NVTANAGDLVVGAGAEINATEGAATLTTSSGKLTTEASHITSAKGVNLSADGGSVAGS 1343
QY 482 IYIDNV 487
Db 1344 INRANV 1349

Search completed: December 19, 2000, 17:02:41
Job time: 1273 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:42:13 ; Search time 248.55 Seconds
(without alignments)
125.108 Million cell updates/sec

Title: US-09-339-159-2

Sequence: 1 LNNFKRIFSTLTLASS.....QSASDSSGGSTIYDNIIVE 490

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 1614 | 61.9 | 516 | 2 JE0134 | mannan endo-1,4-be |
| 2 | 761.5 | 28.2 | 384 | 2 T37232 | secreted beta-mann |
| 3 | 731.5 | 28.1 | 363 | 2 S30386 | mannan endo-1,4-be |
| 4 | 638 | 24.5 | 1331 | 2 A48954 | mannan endo-1,4-be |
| 5 | 427 | 16.4 | 513 | 2 A37219 | mannan endo-1,4-be |
| 6 | 239.5 | 9.2 | 557 | 2 JC5487 | cellulase (EC 3.2. |
| 7 | 234 | 9.0 | 448 | 2 A27631 | cellulase (EC 3.2. |
| 8 | 226.5 | 8.7 | 409 | 2 B25156 | cellulase (EC 3.2. |
| 9 | 226 | 8.7 | 409 | 2 B25156 | cellulase (EC 3.2. |
| 10 | 220.5 | 8.5 | 505 | 2 S39962 | endo-glucanase - Er |
| 11 | 207 | 7.9 | 504 | 2 S54744 | cellulase (EC 3.2. |
| 12 | 203.5 | 7.8 | 570 | 2 S56132 | cellulase (EC 3.2. |
| 13 | 193 | 7.4 | 499 | 2 A27198 | cellulase (EC 3.2. |
| 14 | 189.5 | 7.2 | 508 | 2 A26874 | cellulase (EC 3.2. |
| 15 | 181 | 6.9 | 508 | 2 G69593 | cellulase (EC 3.2. |
| 16 | 180 | 6.9 | 486 | 2 I40548 | cellulase (EC 3.2. |
| 17 | 179 | 6.9 | 466 | 2 C42360 | bifunctional cellu |
| 18 | 173 | 6.6 | 483 | 2 JC5467 | cellulase (EC 3.2. |
| 19 | 171 | 6.6 | 499 | 2 JN0111 | cellulase (EC 3.2. |
| 20 | 170.5 | 6.5 | 584 | 2 JQ1229 | cellulase (EC 3.2. |
| 21 | 169 | 6.5 | 822 | 2 JQ1611 | cellulase (EC 3.2. |
| 22 | 168 | 6.4 | 800 | 2 A29003 | cellulase (EC 3.2. |
| 23 | 166 | 6.4 | 941 | 2 S29043 | cellulase (EC 3.2. |
| 24 | 164.5 | 6.3 | 825 | 2 JS0174 | cellulase (EC 3.2. |
| 25 | 160.5 | 6.2 | 510 | 2 T37541 | probable glycolip |
| 26 | 156 | 6.0 | 611 | 2 JC7177 | endo-glucanase V (E |
| 27 | 155 | 5.9 | 814 | 1 C2C1EM | cellulase (EC 3.2. |
| 28 | 150 | 5.8 | 357 | 2 PC4404 | cellulase (EC 3.2. |
| 29 | 147.5 | 5.7 | 429 | 2 S29044 | endo-glucanase A pr |

ALIGNMENTS

| | | | | | |
|----|-------|-----|------|----------|---------------------|
| 30 | 146.5 | 5.6 | 747 | 2 B47093 | cellulase (EC 3.2. |
| 31 | 140.5 | 5.4 | 426 | 2 A42649 | cellulase (EC 3.2. |
| 32 | 140.5 | 5.4 | 1335 | 2 T17508 | glycoprotein Vp260 |
| 33 | 135 | 5.2 | 438 | 2 A47702 | glucan 1,3-beta-gl |
| 34 | 133 | 5.1 | 1091 | 2 G64964 | hypothetical prote |
| 35 | 132.5 | 5.1 | 1608 | 2 A28182 | hemolysin A - Ser |
| 36 | 132 | 5.1 | 1300 | 2 T00317 | probable serine pr |
| 37 | 131 | 5.0 | 517 | 2 I40798 | cellulase (EC 3.2. |
| 38 | 130 | 5.0 | 515 | 2 S20493 | endo-glucanase - C1 |
| 39 | 129 | 4.9 | 566 | 2 A40589 | cellulase (EC 3.2. |
| 40 | 129 | 4.9 | 1070 | 2 S75712 | cellulase (EC 3.2. |
| 41 | 129 | 4.9 | 1356 | 2 A43855 | high-molecular-wet |
| 42 | 129 | 4.9 | 2020 | 2 C48399 | ABC-type transport |
| 43 | 128.5 | 4.9 | 890 | 2 A30481 | bacteriocin BCNS - |
| 44 | 127 | 4.9 | 32 | 2 PC4278 | quar gum-degrading |
| 45 | 126.5 | 4.9 | 428 | 2 S03767 | cellulase (EC 3.2. |

| | |
|-------------------------------------|-------------------------------------------------------------------------------|
| RESULT | 1 |
| JE0134 | mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans |
| N:Alternate names: | endo-1,4-beta-mannanase |
| C:Species: | Bacillus circulans |
| C:Date: | 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999 |
| C:Accession: | JE0134 |
| R:Yoshida, S.; Sako, Y.; Uchida, A. | Biosci. Biotechnol. Biochem. 62, 514-520, 1998 |
| A:Title: | Cloning, sequence analysis, and expression in Escherichia coli of a gene cod |
| A:Reference number: | JE0134; M01D:98233274 |
| A:Accession: | JE0134 |
| A:Molecule type: | mRNA |
| A:Residues: | 1-516 <YOS> |
| A:Cross-references: | DBJ:AB007123 |
| A:Note: | the authors translated the codon CAA for residue 259 as Asn, CAA for residue |
| A:Note: | the translation of the nucleotide sequence 1294-1365 is not given in this pap |
| C:Keywords: | glycoprotein; glycosidase; hydrolase |

| | | | | |
|-----------------------|-----------------|------------------|-----------|------------|
| Query Match | 61.9% | Score 1614 | DB 2 | Length 516 |
| Best Local Similarity | 59.4% | Pred. No. 1e-102 | | |
| Matches 291 | Conservative 84 | Mismatches 87 | Indels 28 | Gaps 2 |

| | | | |
|----|-----|-----------------------------------------------------------|-----|
| QY | 27 | TSYANSGFYVSGTLYDANGNPFYMRGINGHAWYKDOATTALBGIANTGANTYRIYL | 85 |
| DB | 29 | TNKHAASGFYVSGTKLDATGQPFYMRGVNNAHTYKQSLTAIRPAKTAGNTIRIYL | 88 |
| QY | 87 | SDGGQTKDIDHTNRLISLADNHLVAVPEVHADATGYDSIASLNRAVDYWTMRSALIG | 146 |
| DB | 89 | ANCHKWTLDVVTNNILTECEQNKLIANLEVAHDAGSDSLDANAVYWGIRKALIG | 148 |
| QY | 147 | KEDVTIINANMEFGWEDADWDYKQALPRLRNGLWETLMVDAAGGQCPDSYKNG | 206 |
| DB | 149 | KEDVTIINANMEFGWEDADWDYKQALPRLRNGLWETLMVDAAGGQCPDSYKNG | 208 |
| QY | 207 | REYFNADPQRTNMESTHMYEYAGNAGQVRTNIDRYLNDLALVIGEPHRTNGVDYA | 266 |
| DB | 209 | TEYFNADPQRTNMESTHMYEYAGNAGQVRTNIDRYLNDLALVIGEPHRTNGVDYA | 268 |
| QY | 267 | TIMSYEQRGVGLWAMSKNGPEWEYDLSDWAGNNITLWAGNTLVNPGYGLRSTSL | 326 |
| DB | 269 | TIMSYENRGVGLWAMSKNGPEWEYDLSDWAGNNITLWAGNTLVNPGYGLRSTSL | 328 |
| QY | 327 | TYFTGGSDGSGTSPPT-----LYDEGSGMGQGTSLG | 360 |
| DB | 329 | GTF--GGVTPISPSPTSTPTSTPTPPSPPTSPGNGGNTLYDEGSGMGQGTSLG | 386 |
| QY | 361 | GPNAVTEMSKSHSLKADIQLSNSQHYLAHYONTSLQONSRIQATVHANWGNGM | 420 |
| DB | 387 | GPNAVTEMSKSHSLKADIQLSNSQHYLAHYONTSLQONSRIQATVHANWGNGM | 446 |

QY 421 TATLYKTHGHTWYSGSPFIPNGSSGTTLSLDSNVONLSQVREIGVQPSADSSGOT 480
 DB 447 YATLYKTHGHTWYSGSPFIPNGSSGTTLSLDSNVONLSQVREIGVQPSADSSGOT 506
 QY 481 SITDNVIVE 490
 DB 507 AIVYDVSISLQ 516
 RESULT 2
 T37232
 secreted beta-mannosidase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T37232
 R:Oliver, K.; Harris, D.; Parrhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: 221615
 A:Accession: T37232
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-384 <OLI>
 A:Cross-references: EMBL:AL031514; PIDN:CAA20610.1; GSPDB:GN00070; SCODEB:SC2H4.16
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: manA2; SCODEB:SC2H4.16

Query Match 29.2%; Score 761.5; DB 2; Length 384;
 Best local similarity 44.2%; Pred. No. 19.8e-45;
 Matches 161; Conservative 59; Mismatches 121; Indels 23; Gaps 8;

QY 9 FSTTSLASSILFVSGTSTANANS---GFVYSGTIVDANGNPVPMGINGHAMTKYD 65
 DB 16 FAIVLGLFRA----LAGPSAGRAEAAAGIHVSNGRVVEGNSATVVRVNAATWYTPDR 70
 QY 66 QATTAIEGANTGANTVRIYSDGQMTKDDHTVRNLISLAEDNHLVAVPEVHDATGY- 124
 DB 71 R-TGSTADIADAKANTVRYVLSGGGRMTKSASEVSALLIGCCANKNYICYLEVHDTTGYG 129
 QY 125 --DSIASLNADVWIEMRSALIGKEDVTIINANEMFEGSWEDANADGYKQAIPLRLNA 182
 DB 130 EDAANAATSLDQADADYVWSVKSALEGOEDYVYVNIENEPFGNTNTYATMDATKSAIGKLGA 189
 QY 183 GLNHTLMVDAAGWG--FPOSIDHYGREVNADQRMNMEISHYEYKAGNASQVTRNIDR 241
 DB 190 GLDHALVWDAPNMGQDMSTGRSNAASVFASDDPRTNVTFSVHYTG-VYDTAAEVRDYLNA 248
 QY 242 VLNQDLALVIGFGRHTNGDVDEATIMSYSEORGWGLAMSKNGKPEWEYLDLSNDMA 301
 DB 249 FVSGSLPIVIGFGRDSDNDPDEDAIMATQAQSLGVLGMSNGNGGVYEYIDMNGFID 308
 QY 302 GNNLTAMGNTIVNGPYGLRTSRLSTVF-TGGSDGGSPT-----TLVDFEGSMOG 352
 DB 309 PNLSTLWGNRIFFYGSNGIATSRATATVYGGGGSTGTATANGPYCYVNGASDPDGDGNG 368
 QY 353 WTGS 356
 DB 369 WENS 372
 RESULT 3
 S30386
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) precursor - Streptomyces lividans
 N:Alternate names: beta-mannanase
 C:Species: Streptomyces lividans
 C:Date: 02-Dec-1993 #sequence_revision 13-Mar-1997 #text_change 22-Oct-1999
 C:Accession: S30386; S27699
 R:Archand, N.; Kluepfel, D.; Paradis, F.W.; Morosoli, R.; Shareck, F.
 Biochem. J. 290, 857-863, 1993
 A:Title: beta-Mannanase of Streptomyces lividans 66: cloning and DNA sequence of the man
 A:Reference number: S30386; MUID:93207541
 A:Accession: S30386

A:Molecule type: DNA
 A:Residues: 1-363 <ARC>
 A:Cross-references: EMBL:M92297; NID:g153193; PIDN:AAA26710.1; PID:g404076
 A:Experimental source: strain 1326
 C:Genetics:
 A:Gene: manA
 C:Function:
 A:Description: catalyzes degradation of beta-mannans into short-chain oligo-mannoside
 A:Pathway: mannose metabolism
 A>Note: Industrially used for bleaching of wood pulps
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F.1.36/Doman: signal sequence #status predicted <SIG>
 F.37-363/Product: mannan endo-1,4-beta-mannosidase #status predicted <MAT>

Query Match 28.1%; Score 731.5; DB 2; Length 363;
 Best local similarity 46.1%; Pred. No. 1e-42;
 Matches 155; Conservative 55; Mismatches 111; Indels 15; Gaps 8;

QY 13 LSTLLASSILFVSGTSTANANS---GFVYSGTIVDANGNPVPMGINGHAMTKYD 69
 DB 19 LGILFRA----LAGPSAGRAEAAAGIHVSNGRVVEGNSATVVRVNAATWYTPDR-TG 72
 QY 70 ATEGIANGTANTVRIYSDGQMTKDDHTVRNLISLAEDNHLVAVPEVHDATGY--DSI 127
 DB 73 STADIADAKANTVRYVLSGGGRMTKSASEVSALLIGCCANKNYICYLEVHDTTGYGKGA 132
 QY 128 ASINRAVDWIEMRSAL-IGKEDVTIINANEMFEGSWEDANADGYKQAIPLRLNAGLNH 186
 DB 133 TSDIQAGDVIYVGSKAMRAQEDYVYVNIENEPFGNTNTYATMDATKSAIGLRGGLGH 192
 QY 187 TLMVDAAGWG--FPOSIDHYGREVNADQRMNMEISHYEYKAGNASQVTRNIDRLNQ 245
 DB 193 ALMWDAPNMGQDMSTGRSNAASVFASDDPRTNVTFSVHYTG-VYDTAAEVRDYLNAFVN 251
 QY 246 DLALVIGFGRHTNGDVDEATIMSYSEORGWGLAMSKNGKPEWEYLDLSNDMAGNNL 305
 DB 252 GLPIVIGFGRDSDNDPDEDAIMATQAQSLGVLGMSNGNGGVYEYIDMNGFDPNL 311
 QY 306 TANGNTIVNGPYGLRTSRLSTVF-TGGSDGGSPT 340
 DB 312 TSWGNRIYGSNGIATSRATATVYGGGGSTGTATAP 347

RESULT 4
 A48954
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
 N:Alternate names: beta-mannanase
 C:Species: Caldocellum saccharolyticum
 C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
 C:Accession: A48954; A43745
 R:Idiba, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
 Appl. Environ. Microbiol. 58, 3864-3867, 1992
 A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a man
 A:Reference number: A48954; MUID:93119139
 A:Accession: A48954
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-1331 <GIB>
 A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
 A>Note: sequence extracted from NCBI backbone (NCBI:121576; NCBI:121577)
 R:Luethl, E.; Bhana Jaemai, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
 Appl. Environ. Microbiol. 57, 694-700, 1991
 A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene
 A:Reference number: A43745; MUID:91247819
 A:Accession: A43745
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-337; 'P'ROHORO' <LUE>
 A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
 A>Note: the authors translated the codon CAC for residue 262 as Glu
 A>Note: this sequence has been revised in reference A48954
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

[illegible]

F:1-29/Domain: signal sequence #status Predicted <SIG>
F:30-557/Product: cellulase #status Predicted <MAF>

Query Match 9.2%; Score 239.5; DB 2; Length 557;
Best Local Similarity 23.1%; Pred. No. 6.7e-09;
Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

```

OY 2 NNGFKRIFSI-----TLSSLLASSILFVSGTS-----TANAN-----SGFTV 38
DB 3 NNVKKIILSIYVAGAMLMALVPINVALETYSNLGNNAKKPSYVGKQLQILNNKI-- 60
OY 39 SCFTIYDANGNPFVNGIN-HGHAWTKDQATTAIEGIANTGA-----NTVATV- 86
DB 61 --KTLCDKGNPQLQKMGSTHGLQW-----PGVANNNAFALNSDWNNSVIRLAMY 110
OY 87 -SDGQWTDODI-HYVRNLSLAEDNHLVAVPEVHDATGYDSTASLNRAVDYWIEMRSAL 144
DB 111 VAEQGYATNPVKQYVYININATANDMVYIDWHMANNPDPNASTYSGAQSFNDISTYL 170
OY 145 ICKEDVIINIANEMFGSWEG---DA--WADGVKQALP---HLRNAGLHHTLMVDAAGM 196
DB 171 YNNKNRIITELCNERPENGCVTNDATGMAQYKSTATPIVQLRNKGNENLITVGPFWS 230
OY 197 QPPOSIDHYGREVFADPQRTMFSIMHYEAGN-ASQVRTNIDR-----VLNODIA 248
DB 231 QRPDLAAD-----NPINDNTMYSVHF--YSGTNPISVVDNBRDAMSNVRYALNHGA 282
OY 249 LVIGEFGHHTNGVDENATIMSYSEQRGVGLAMSKMGKPEWEYLDLSND----- 299
DB 283 VYATWMTGSLATG---TGPLYL-AKADAWLDF-LNGNNISMCFEJSINKDEKAALNS 335
OY 300 -----WAGNLPZAMG-----NTIVNGPYGLRSTRSLSTVETG 332
DB 336 LMSLDPGSKMLADMBELTSSGOYVARIRKGAATYATPVDTYQTPAPKDFSSGFDFENDG 395
OY 333 GSDG---GTSPTIYDEFGSMQGTGSLSGPMAVTEMSSKGSLSK---ADIOLS 383
DB 336 TTQGFVNDPDSPTITAINVENNANKISNL-----NKSQNDLSEGNFMANVRIS 445
OY 384 SNS-OHYLVHVIQNTSLQONSRIQATYKHNMGSVGNGMTARLYVTGHR-----TW 434
DB 446 ADIWQOSINITYGDKITMDVYIAPTPV-----NVSITAIIPQSTHGMNPTAIRW 496
OY 435 YSGSEVPINGSSGTTSLDLSNVQNLQVREIGVQFOSASDS-----SGQSTIY 484
DB 497 TNNFVAQDGTATKATLTISTNDSNPNFTIA-----TDAADEVYNNMLFVGSNDNISL 550
OY 485 DNV 487
DB 551 DNI 553

```

RESULT 7

A27631
cellulase (EC 3.2.1.4) precursor - Clostridium acetobutylicum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium acetobutylicum
C>Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum P2
A:Reference number: A27631; MUID:88268074
A:Accession: A27631
A:Molecule type: DNA
A:Residues: 1-448 <ZAP>
A:Cross-references: EMBL:M31311; NID:q144789; PIDN:AAA23230.1; PID:q144790
A>Note: the authors translated the codon GAG for residue 116 as Gly, GAA for residue 263
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
lulose
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 9.0%; Score 234; DB 2; Length 448;
Best Local Similarity 22.6%; Pred. No. 1.2e-08;
Matches 115; Conservative 75; Mismatches 185; Indels 134; Gaps 24;

```

OY 3 NNGFKRIFSI-----TLSSLLASSILF-VSGSTANANSG-----FYVSGTLYDANGNPFVNR 54
DB 9 NFFKTFEFLAVVA---MLFTVLSTNTYKAAPATTSGGQLKVVGSQDCSNGKPIQLK 65
OY 55 GI-NHGHMY-----KDQATTAIEGIANTGANTRYIVL--SDGQWTD--KDDI 97
DB 66 GMSHGLQWYVYVYDMSKFLRDK-----GQVNVIRAMYTEGYSNPSSQK 115
OY 98 HYVRNLSLAEDNHLVAVPEVHDATGYDSTASLNRAVDYWIEMRSALIGKEDTYINIAN 157
DB 116 EKIKRIVQDADILNMYIIDHILSDNNPNTYKQAKSFPEMAEE-YGKTSNVIYEICN 174
OY 158 EWFQSWEDDANADYK---QAIPELRNAGLHHTLMVDAAGMGPPOSIDHYGREVFAD 213
DB 175 EPNG---GTNNANDIKPYANTYIIPAIADIPNNIITVGTSTWSQVDIAD-----N 223
OY 214 FOR--NTFESTIMTEYAGNNSQVKTNIDRYLNODLALVIGFGRHRTNGVDENATIMSY 271
DB 224 PLRYSNIMYTFHF--YAGTHQSLRDKINYAMSKGIAIFTEWGTSDASGN----- 272
OY 272 SEQRGVGLAMSKMGKPEWEYLDLSNDW---AGNNTLNGCNTIVNGPYGLRSTRSLST 327
DB 273 -----GGP---YLDESQRWVDFMAKSNL-SWTN-----WALCDKSEASA 307
OY 328 VFTGGSDGTSPTIYDEFGSMQGTGSSLS-----GGPMVAVTEMSSKGSLSK 377
DB 308 ALKSSGRTG-----GWDSDLTITGLFVKKSIGG-----SNTTSQISA 346
OY 378 ADIOLSSNQYHLVHVIQNTSLQONSRIQATYKHNMGSVGNGMTARLYVK--TGHR 434
DB 347 PTFSLQSGTYSAQTVLTSSDNDSDVIHYTDTGPTTSSPYVTSPIRISKTYVKAFTT 406
OY 435 YSGSEVPINGSSGTTSLDLSNVQNLQV 463
DB 407 KTG---MTDSNITSVAYTISNTDPAKV 431

```

RESULT 8

A25156
cellulase (EC 3.2.1.4) 1 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. s
A:Reference number: A91825; MUID:87056924
A:Molecule type: DNA
A:Residues: 1-488 <FUK>
A:Cross-references: GB:M44781; GB:X53449; NID:q142659; PIDN:AAA22301.1; PID:q142660
A:Experimental source: strain N-4, plasmid pNK1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
cellulose
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 8.7%; Score 226.5; DB 2; Length 488;
Best Local Similarity 23.1%; Pred. No. 4.3e-08;
Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;

```

OY 6 KKSISLSSLLASSILFVSGSTRAN-----ANSGRYSGTLYDANGNPFVNRGI-NH 58
DB 2 KKLITLITVFTLA--LLEVGSTSNANGSVVEQGLSQNGQLVNHGDPVQLKGMSSH 59

```

```

OY 59 GAAWTKDAQATALEGIANTGANTVRIVLSDGGQWTKD-----IHE----- 99
| 11
DB 60 GLOWTG-----QFVNTYDSIKMLRDDMGITVFRAMATSSGGCTIEDPS 101
OY 100 ---VRNLISLAEDNHLVAVPEVDATGYDSIASLRNAVDTIEMRSALIGKEDTYIINI 155
| 11
DB 102 VAEKKEAVEAALDGLIYIIMHILSDNDPNYKKEAEFEDEM-SALYGDYPNVIYEI 160
OY 156 ANEMFG---SWEGDAMADGY-KQALPRLNAGINHTLMTVDAAGWGQFPOSINDYGREVEN 211
| 11
DB 161 ANEPNGHNVRM--DSHIKRYAEVLPVIRANDPNNIIVIGTATWS---ODVHEADNQDL 215
OY 212 ADPQRTMFSIMAYEYAGNASOVRTNIDRLVNOALVIGEGHHTNGD-----VDENF 267
| 11
DB 216 -DP--NMAIAHFH--YAGTHGOQLRQVDYALSRGALITVSEMGTSAAATGDGVLDEAQ 270
OY 268 I-MSYSEQNGVGLWMS-----WKGNGPEWEYIDLSDNMAGNMLTANGNTIVNG 315
| 11
DB 271 VWIDPMDERNLSWAMWSLTHKDESSAALMPGANP-----TGWTAALSLFSGAFVRE- 322
OY 316 PYGLRETSRLSTVFTGGSD---GGTSPTTLTYDFEGSMQGWGSSL-----SGGPWA 364
| 11
DB 323 --KIKESASIPPSDPTPPSDPDGEPDPPTPSD-PGEYPAAMPNQIYTNIEIYHNGQLMQ 379
OY 365 VTEWS 369
| 11
DB 380 AKWMT 384

```

RESULT 9
B25156
cellulase (EC 3.2.1.4) 2 - *Bacillus* sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: *Bacillus* sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: B25156
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain
A:Reference number: A91825; MUID:87056924
A:Accession: B25156
A:Molecule type: DNA
A:Residues: 1-409 <FUNK>
A:Cross-references: GB:M4729; NID:g142655; PIDN:AA22299.1; PID:g142656
A:Experimental source: strain N-4, plasmid pMK2
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

```

Query Match 8.7%; Score 226; DB 2; Length 409;
Best Local Similarity 23.2%; Pred. No. 3.6e-08;
Matches 104; Conservative 62; Mismatches 147; Indels 136; Gaps 24;
OY 6 KRIFFITSLILASSILFVSGTSTA-----NANSGFVSGTTLTDANGNFPWARGT-N 57
| 11
DB 2 KRITTFVLLMTTLA-LFTIGMTAADVSYVEEGQLSISNGELVNDGEPVQLKMS 60
OY 58 HGHAAYKQOATFAEGIANGTANTVRIYVLSGGQWTKD-----IHT----- 99
| 11
DB 61 HGLQWYG-----QFVYESMKMLRDDMGITVFRAMATSSGGCTIEDP 102
OY 100 ---VRNLISLAEDNHLVAVPEVDATGYDSIASLRNAVDTIEMRSALIGKEDTYIINI 154
| 11
DB 103 SVKEKKEAVEAALDGLIYIIMHILSDNDPNYKKEAEFEDEM-SELYGDYPNVIYE 161
OY 155 IANEMFGS---WEGD--AMADGYKQALPRLNAGINHTLMTVDAAGWGQFPOSINDYGREV 209
| 11
DB 162 IANEPNGSVTVNDNQIKPFAE---EVITVPIRNDPNNIIV--GTGTWSQDVFH-AAQN 214
OY 210 FNADPQRTMFSIMAYEYAGNASOVRTNIDRLVNOALVIGEGHHTNGD-----VDE 265
| 11

```

```

DB 215 QLTLD--NMYAFAHF--YAGTHGOQLRQVDYALDOGAALFVSEMGTSBATGDGVLDE 270
OY 266 ATI-MSYSEQRGVGLWMS-----WKGNGPEWEYIDLSDNMAGNMLTANGNTIV 313
| 11
DB 271 AQVWIDPMDERNLSWAMWSLTHKDESSAALMPGASP-----TGWTAALSLFSGTFV- 322
OY 314 NPGYGLRETSRLSTVFTGG-----SDGFI-----SPTTLTD 345
| 11
DB 323 ---REKIRRESATTPPSDPTPPSDPDGEPDPPTPSD-PGEYPAAMPNQIYTNIEIYHNGQLMQ 379
OY 346 FE-----GSK---QGWGSSLG---GPW 363
| 11
DB 377 DEIYVHNGQLRQVDAKMTQNOEPGDPYRGPW 405

```

RESULT 10
S39962
endoglucanase - *Erythrina carotovora*
C:Species: *Erythrina carotovora*
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39962
R:Cooper, V.J.C.; Salmond, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A:Title: Molecular analysis of the major cellulase (CelV) of *Erythrina carotovora*: evid
A:Reference number: S39962; MUID:94067016
A:Accession: S39962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <COO>
A:Cross-references: EMBL:X76000; NID:g434941; PIDN:CA53592.1; PID:g434942

```

Query Match 8.5%; Score 220.5; DB 2; Length 505;
Best Local Similarity 21.8%; Pred. No. 1.2e-07;
Matches 116; Conservative 69; Mismatches 209; Indels 137; Gaps 21;

```

```

OY 2 NNGFKRISITISLILASSILFVSGTSTRANNSG-FYVSGTTLTYANGNFPWARGT-NHG 59
| 11
DB 6 NOIVKRLTGAVTYVGLSLSFSAALNPVETHGQLSIENGRVDEQGRVQLRISHS 65
OY 60 HAWYKDAQATALEGIANTGANTVRIVLSDGGQWTKD----- 96
| 11
DB 66 LQWFGD-----YKNDMSKMLRDDMGITVFRAMATADGYISNPSL 107
OY 97 IHTVNLISLAEDNHLVAVPEVDATGYDSIASLRNAVDTIEMRSALIGKEDTYIINI 156
| 11
DB 108 ANKREAVEAALDGLIYIIMHILSDNDPNYKQATFEFEM-AGLYGSSPNVIYEI 166
OY 157 NEMFG--SWEGD--AMADGYKQALPRLNAGINHTLMTVDAAGWGQFPOSINDYGREVEN 212
| 11
DB 167 NEPNGVWNGQIRPYA---LEVDTIRSKDPDNIIV--GTGTWSQDIHD--AADNQ 217
OY 213 DPQRTMFSIMAYEYAGNASOVRTNIDRLVNOALVIGEGHHTNGD-----VDENF 267
| 11
DB 218 LPDPTMTALHF--YAGTHGOQLRQVDYALSRGALITVSEMGTSAAATGDGVLDEAQ 270
OY 268 IMSYSEQRGVGLWMSWKGNGPEWEY---DLSNDMAGNMLTANGNTIVNGPYGLRETS 323
| 11
DB 276 WIDFLNNGGVSWNMSLTDKSEASALABGASKSGGWTEQNLSGKFV-----REQI 328
OY 324 RLSTFTTGGSGSGTSPPTLTDPEBSMGTGSSLSGGPNAVTEWSKSGSHSLKADIQLS 363
| 11
DB 329 R-----AGANLGGGDTPTTPEPTPNNGTGT-----DVLV- 359
OY 384 SNSQHYLHYQTSLOQNSRIQATVKHANWGSV---NGMTRLYVKTGHGTYWSSGFV 440
| 11
DB 360 ---QRYVNDNPS---DDALPMAVINIKTGSTPKLSLDLQVRYFPHD----- 400
OY 441 PINGSGTTLSDLSNVQNLGSOVREIGVQFOSASD-----SSGQTSI 482
| 11
DB 401 -DGRPGANLFFYDMANVGPNNIYVITGTTPAASDANKRNVYLVTFSSGAGSL 449

```

RESULT 11

cellulase (EC 3.2.1.4) precursor - *Erynia carotovora* (SCC 3193)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Erynia carotovora*

A:Variety: SCC 3193

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S54744; S44996

R:Maee, A.; Heikkinen, R.; Palva, E.T.

Mol. Gen. Genet. 247, 17-26, 1995

A:Title: Structure and regulation of the *Erynia carotovora* subspecies *carotovora* SCC3193

A:Reference number: S54744; M01D:95231512

A:Accession type: DNA

A:Residues: 1-504 <MAE>

C:Genetics: EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g493493

A:Gene: celV1

C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-39/Domain: signal sequence #status predicted <SIG>

F:33-504/Product: cellulase #status predicted <MAT>

Query Match 7.8%; Score 203.5; DB 2; Length 570;

Best Local Similarity 23.3%; Pred. No. 26-06; Mismatches 197; Indels 119; Gaps 26;

Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

14 SILLASSILFVSGSTANANGFYVSGTLLYDANGNPFVNGIN-----HGHWYKD 65

DB 22 AILLGSGGLGVAQADVAFLSVQGNKIL-ANGQAFSGMSLFMSNTMGKRYNA 80

QY 66 QATTAIEGANTGANTVIVL---SDGGQWT---KODIHVYRNLSIAEDNHLVAPEV 118

DB 81 QVSWLK--SPMNAVLPAAGVDEGGYLPDPANKQVTOYVD-AAIANDMYIIMHS 137

QY 119 HDATGYDSIASINRAVDYIEMRSALIGKEDTVIINANWFG-SWEGDANADGKQA-I 176

DB 138 HNAHQYQS---QALAFQEK-ARKYGANNHVIYEINEPLQVSMWT--IKPYAQAVI 189

QY 177 PRLNAGNLHTLVADDAWGQFPOSIDHYGREYRNPAD---QRNTMESIHMEYAGNAS 233

DB 190 AAIKRAIDPNDLITVGTPTWSQ-----DVDVANDITGYONTAYTLHF--YAGTGO 239

QY 234 QYRTNIDRYLNDLALVIGEGHRTNGDVDEATIMSYSEORGVGLAMSKNGPEWEY 293

DB 240 YLRDRAQRLNGLALFVTEMGSVANED---GAVANSE-----GW 315

QY 294 LDLSNDMAGNNLTAMGNTIVNGPYGLRETSLVFTGGSDGTSPTLLYDFEGSMQGW 353

DB 276 ---TNAVSEFKT---NHISNANMALNDKVEGASALVPGASANG-----GW 315

QY 354 TQSSSL-SGGPFA---VTEM---SSKGSLSLKADQLQSLNSGHIHYLHYONTLSLOONSRHQ 405

DB 316 VNSQLTASGALAKSITISGPPSYNTSSSAVSSQTOVSSSO--APVSSSSSTKSSVVS 373

QY 406 ATYV--KHANWGSVNGMNTARLYVKTGHGYTW-----YSGSFYPINGSSGTTLSL 452

DB 374 SAVSGQOCWV---YGLTLPLOCSTTNGMGWENNSCIARATCSQAPAPWGLVGSSTSS- 428

QY 453 DLSNVONLSQVREIGVQFOSADSSGQTSI 482

DB 429 -----QASSSVRSSSSSLVSSRSSSSSV 453

RESULT 12

cellulase (EC 3.2.1.4) precursor - *Pseudomonas fluorescens*

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Pseudomonas fluorescens*

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

RESULT 13

cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain IF03034)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus subtilis*

C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1995

R:Nakamura, A.; Uozumi, T.; Beppu, T.
 Eur. J. Biochem. 164, 317-320, 1987
 A:Title: Nucleotide sequence of a cellulase gene of *Bacillus subtilis*.
 A:Reference number: A27198; MID:87190397
 A:Accession: A27198
 A:Molecule type: DNA
 A:Residues: 1499 <NAX>
 A:Cross-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671
 A:Experimental source: strain IF03034
 A:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 C:Keywords: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 8,5e-06;

Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

```

QY 6 KRTFTSLSLASSILF-----VSGTSTANNSG-FYVSGTTLTDANGNPFYMRG 55
DB 2 KRISITITLLAVLTMGGLSPASAGCTKTPAAKNGQLSITQTLVNRDRAVOLKG 61
QY 56 I-NHGAWKYDQATTAEGTANTGANTVRLVSDGGQWTKD-----IHT----- 99
DB 62 ISSHGLQWTD-----FVNKDSLKLWLDMDGIVFRAMATYADGGYI 103
QY 100 -----VANLSLADNHLVAPPEVHDATGDSIASLNRADVNIEMRSALIGKEDY 151
DB 104 DNPVKNKRYEAEVAKELGIYIIDWHLNDGNPNQNEKRADEFKEM-SSLYGTPNV 162
QY 152 IINANEMFG--SWEQD--AMADGYKQALPRLNAGLHNTLAMDAGMGQFPOSINDYGR 207
DB 163 IYELANEPNDGVNKKRDKRYAE---EVSIVIRKNDPDNIIV---GTGWSQDQVDN--- 213
QY 208 EVENADPQ---RNTMFSIHNEYAGNAGSQRNTIDRVLMQDLALVIGEGHRTNGD-- 262
DB 214 ---AADDQLDANVMYALH--YAGTHGQSLRQKANYALSKAPITFTEWGTSDASGNG 268
QY 263 --VDEATIMSSEORGVGLWMSKNGNPEM--EYLDLSN--DMAGNNLTANGNTIVNGPYG 318
DB 269 VFIDQSR-----EMLNTLDSKNSISWVNNL----- 293
QY 319 LRETSRLSTVTFGGSGDGTPTLTDPEGSKOGWTGSSLSGGPMAVETSSKGSLSKA 378
DB 294 -----SDKQSSSALK-----PGASKTGC-WPLTDLTASGTF-VRE 327
QY 379 DI--QLSSNQHYLVHVIQNTSLQONSRIQATVTHANNGSVNGMTALVYKTHGTY--- 433
DB 328 NILGNKDSIKERPEPQADNPADENG-ISOYTRAGDGVNSNQTLPQLHIKNGMATVDL 386
QY 434 -----WYSGFVPIINGSGTTLSDLSNV 457
DB 387 KDVTARYWYNAK-----NKGQNFDCDYAQI 411

```

RESULT 14

A26874
 cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain DLG)
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 08-Mar-1989 #sequence, revision 08-Mar-1989 #text, change 15-Oct-1999
 C:Accession: A26874; B26874
 R:Robson, L.M.; Chambliss, G.R.
 J. Bacteriol. 169, 2017-2025, 1987
 A:Title: Endo-beta-1,4-glucanase gene of *Bacillus subtilis* DLG.
 A:Reference number: A26874; MID:87194581
 A:Accession: A26874
 A:Molecule type: DNA
 A:Residues: 1-508 <ROB2>
 A:Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
 A:Experimental source: strain DLG

A:Accession: B26874
 A:Molecule type: protein
 A:Residues: 39-53 <ROB2>
 A:Experimental source: strain DLG
 A:Note: the authors believe Met-1 and Met-2 may be alternate initiators
 C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; po
 F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 7.2%; Score 188.5; DB 2; Length 508;
 Best Local Similarity 21.1%; Pred. No. 1.8e-05;

Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17

```

QY 6 KRTFTSLSLASSILF-----VSGTSTANNSG-FYVSGTTLTDANGNPFYMRG 55
DB 11 KRISITITLLAVLTMGGLSPASAGCTKTPAAKNGQLSITQTLVNRDRAVOLKG 70
QY 56 I-NHGAWKYDQATTAEGTANTGANTVRLVSDGGQWTKD-----IHT----- 99
DB 71 ISSHGLQWTD-----FVNKDSLKLWLDMDGIVFRAMATYADGGYI 112
QY 100 -----VANLSLADNHLVAPPEVHDATGDSIASLNRADVNIEMRSALIGKEDY 151
DB 113 DNPVKNKRYEAEVAKELGIYIIDWHLNDGNPNQNEKRADEFKEM-SSLYGTPNV 171
QY 152 IINANEMFG--SWEQD--AMADGYKQALPRLNAGLHNTLAMDAGMGQFPOSINDYGR 207
DB 172 IYELANEPNDGVNKKRDKRYAE---EVSIVIRKNDPDNIIV---GTGWSQDQVDN--- 222
QY 208 EVENADPQ---RNTMFSIHNEYAGNAGSQRNTIDRVLMQDLALVIGEGHRTNGD-- 262
DB 223 ---AADDQLDANVMYALH--YAGTHGQSLRQKANYALSKAPITFTEWGTSDASGNG 277
QY 263 --VDEA-TIMSSEORGVGLWMSKNGNPEM---DLSMDAGNNLTANGNTIVNG 315
DB 278 VFIDQSR-----EMLNTLDSKNSISWVNNLSDKQSSSALKPGASKTGWPLTDLTASGTFVREN 337
QY 316 PYGLRERS 324
DB 338 IRGTKDSIK 346

```

RESULT 15

G69593
 cellulase (EC 3.2.1.4) bglc precursor - *Bacillus subtilis*
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence, revision 05-Dec-1997 #text, change 15-Oct-1999
 C:Accession: G69593; A26114; I40353; S24233; S49103; I39803
 R:Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Bruchl, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emmerich, P.T.; Entian, K.D.; Errington, J.; Fabbre, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henatt, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani
 A:Authors: Schleich, S.; Schwoerer, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se
 akuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumberg, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MID:98044033
 A:Accession: G69593
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-508 <KUN>

A:Cross-references: GB:299113; GB:AL009126; NID:92634090; PIDN:CAB13696.1; PID:el183471.
 A:Experimental source: strain 168
 R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moranello, F.; Seld
 Nucleic Acids Res. 14, 9159-9170, 1986
 A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
 A:Reference number: A26114; MOID:87066783
 A:Accession: A26114
 A:Molecule type: DNA
 A:Residues: 10-508 <MAC>
 A:Experimental source: strain PAP115
 A:Note: part of this sequence, including the amino end of the mature form, was confirmed
 R:Lindahl, V.; Ae, K.; Tromso, A.
 Antonie Van Leeuwenhoek 66, 327-332, 1994
 A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C
 A:Reference number: I40353; MOID:95225656
 A:Accession: I40353
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-291 'N', 293-508 <LIN2>
 A:Cross-references: EMBL:X67044; NID:939776; PIDN:CAA47429.1; PID:939777
 R:Lindahl, V.; Ae, K.
 A:Experimental source: strain CK-2
 R:Wolf, M.; Geetzl, A.; Borries, R.
 submitted to the EMBL Data Library, December 1993
 A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons
 A:Reference number: S49103
 A:Accession: S49103
 A:Molecule type: DNA
 A:Residues: 10-508 <MO>
 A:Cross-references: EMBL:Z29076; NID:9509266; PIDN:CAA82317.1; PID:9509267
 R:Seo, Y.S.; Lee, Y.H.; Park, U.H.; Kang, H.
 Korean J. Microbiol. 24, 236-242, 1986
 A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
 A:Reference number: I39803
 A:Accession: I39803
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'F', 23, 'A', 25-26, 'A', 29-31, 'P', 33, 'PQ', 36-
 A:Cross-references: GB:M38634; NID:9142657; PIDN:AAA22300.1; PID:9142658
 A:Experimental source: strain ATCC 6633
 A:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
 C:Genetics:
 A:Gene: bglC
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:39-508/Product: cellulase #status predicted <MAT>

Query Match 6.9%; Score 181; DB 2; Length 508;
 Best local similarity 19.3%; Pred. No. 5.8e-05;
 Matches 100; Conservative 102; Mismatches 188; Indels 128; Gaps 26;

OY 6 KRTFTSLTSLAASSLIF-----VSGSTANANSQ-FYVSGTLLDANGNPFVMBG 55
 DB 11 KRISISIFITCLLTLLTGMGMIASPASAGTKTPVAKNGOLSIKQTOLVNRDOKAVQLKG 70
 OY 56 I-NHGAHWYKQAT-TAIEGIANAGANTV---RIVLSDGGQWTKDDI-HTVNLISLAD 109
 DB 71 ISSHGLQWYGEYVKNKSLKRLRDMGWITFRAMYTADGGYIDNPSYKKNVKEAVEAKE 130
 OY 110 NHLVAPVEDATGYDSIASLNRADVWIMRSGALIGKEDTYLINTANEMFG--SWEGD- 166
 DB 131 LGIYVLDHMLNDGNPNOKRERAKREFKEM-SSLYGNIDPNVIEIANEPNGVDYMKRDI 189

OY 167 -AWADGYKQALPRLRNAGLNTLWDAAGWGPPOSTIHDIYGREVFNDPO---RNTMPSI 222
 DB 190 KPYAE---EYISVIRKNDPDNIIV---GTGTWSODVND-----AADDLKDNANWYAL 237
 OY 223 HMEYAGNNAQVATNIDRYLNQDLALVIEFGHRTNGD---VDPA-TIMSGSEORGV 277
 DB 238 HF--YAGTHGQFLDKANAYLSKAPIFVTEWGTSDASGNGVFLDOSREMLKYLDSKTI 295
 OY 278 GWLAWSMKNGPWEWEYL---DLSNDWAGNNLTAMGNTVNGPYGLRETSR----- 324
 DB 296 SWVMNNLSDKQESSALKPGASKTGWRSLDLSASGTFVRENILIGTIDSTKIDIPETPSKD 355
 OY 325 -----ISTYFTGGSGSGSPPTLYDFEBSMGWGTGSSLSGPFMAVTESSKSHSLK 377
 DB 356 KPTQENGISVYVRAG-----DGSN-----NSNQIR 380
 OY 378 ADIOSSN-----SQHYLVIONTSIQQNSRIQATVKHANMGVSGNGMTARLYV- 426
 DB 381 POLQIKNNGNNTYDLKQVYATRYWKANKNG--QNF-----DDDYAQIGCGNVTYHKEFV 437
 OY 427 --KTGHGTYWY-----SGSFVPINGSSGTTLSLDSL 456
 DB 432 LKPKQAGDYLELGFKNQTLAP--GASTGNIQRLHN 467

Search completed: December 19, 2000, 17:06:42
 Job time: 1469 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:19:35 ; Search time 132.02 Seconds

(without alignments)
118.577 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNGFKKISITSLILASS.....QASDSSGQSIYIDNVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 737.5 | 28.3 | 383 | 1 MANA_STRL1 | P51529 streptomyc |
| 2 | 638 | 24.5 | 1331 | 1 MANB_CALSA | P22533 caldocellu |
| 3 | 427 | 16.4 | 513 | 1 MANB_BACSM | P16699 bacillus sp |
| 4 | 234 | 9.0 | 448 | 1 GUN_CLOAB | P15704 clostridium |
| 5 | 229.5 | 8.8 | 400 | 1 GUN5_BACAG | O85465 bacillus ag |
| 6 | 226.5 | 8.7 | 488 | 1 GUN1_BACSA | P06566 bacillus sp |
| 7 | 226 | 8.7 | 409 | 1 GUN2_BACSA | P06565 bacillus sp |
| 8 | 220.5 | 8.5 | 505 | 1 GUNV_ERMCA | O47096 erwina car |
| 9 | 219.5 | 8.4 | 444 | 1 GUNV_ERMCA | O59394 erwina car |
| 10 | 207 | 7.9 | 504 | 1 GUNW_ERMCA | O59395 erwina car |
| 11 | 188.5 | 7.2 | 499 | 1 GUN1_BACSU | P07983 bacillus su |
| 12 | 181 | 6.9 | 499 | 1 GUN2_BACSU | P10475 bacillus su |
| 13 | 180 | 6.9 | 466 | 1 GUN5_THERU | O01786 thermomon |
| 14 | 171 | 6.6 | 499 | 1 GUN3_BACSU | P23549 bacillus su |
| 15 | 170.5 | 6.5 | 584 | 1 GUN2_CIOCE | P25472 clostridium |
| 16 | 168 | 6.4 | 800 | 1 GUN_BACSL | P06564 bacillus sp |
| 17 | 166 | 6.4 | 941 | 1 GUN_BACSL | P06564 bacillus sp |
| 18 | 165.5 | 6.3 | 459 | 1 GUN3_STRL1 | P21035 streptomyc |
| 19 | 164.5 | 6.3 | 825 | 1 GUN3_BACSA | P19570 bacillus sp |
| 20 | 160.5 | 6.2 | 510 | 1 YDYL_SCHPO | O13692 clostrid |
| 21 | 155 | 5.9 | 814 | 1 GUN2_CIOCE | P10472 clostridium |
| 22 | 151.5 | 5.8 | 562 | 1 GUN1_ACTIE | P55583 actinobacter |
| 23 | 147.5 | 5.7 | 429 | 1 GUN4_BUTFI | P22541 butyrivibri |
| 24 | 146.5 | 5.6 | 747 | 1 GUN4_CELFI | P25400 cellulomona |
| 25 | 141 | 5.4 | 312 | 1 GUN4_RUMAL | O07940 ruminooccu |
| 26 | 140.5 | 5.4 | 426 | 1 GUN4_BURSO | P19794 burkholderi |
| 27 | 133 | 5.1 | 1039 | 1 AG43_ECOCI | P39180 escherichia |
| 28 | 132.5 | 5.1 | 1608 | 1 HXKA_SERVA | P13320 serraria ma |
| 29 | 132 | 5.1 | 438 | 1 EXG_CANAL | P28717 candida alb |
| 30 | 131 | 5.0 | 515 | 1 GUN2_CIOCE | P28623 clostridium |
| 31 | 131 | 5.0 | 517 | 1 GUN4_CIOCE | P54937 clostridium |
| 32 | 129 | 4.9 | 566 | 1 GUN3_CIOCE | O05332 clostridium |
| 33 | 129 | 4.9 | 2003 | 1 YDAB_ECOCI | P33666 escherichia |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 128.5 | 4.9 | 890 | 1 BCN5_CLOPE | P08696 clostridium |
| 35 | 127 | 4.9 | 426 | 1 GUN2_ERMCH | P07103 erwina chr |
| 36 | 123.5 | 4.7 | 645 | 1 LIP1_PROHU | P40601 photorhabdu |
| 37 | 123.5 | 4.7 | 711 | 1 CDGT_BACST | P31797 bacillus st |
| 38 | 122.5 | 4.7 | 1039 | 1 GUNB_CALSA | P10474 c endogluca |
| 39 | 118.5 | 4.5 | 1829 | 1 FRPC_NEIMC | P5127 neisseria m |
| 40 | 116.5 | 4.5 | 440 | 1 GUNB_CIOCE | P28623 clostridium |
| 41 | 116.5 | 4.5 | 976 | 1 RIBP_ADEB3 | O03553 bovine aden |
| 42 | 115 | 4.4 | 1656 | 1 OMPB_RICCA | O06653 r outer mem |
| 43 | 114.5 | 4.4 | 865 | 1 HTRF_ECOCI | P33129 escherichia |
| 44 | 114 | 4.4 | 802 | 1 PEFC_SALTY | P37868 salmonella |
| 45 | 114 | 4.4 | 2334 | 1 WAPA_BACSU | O07833 bacillus su |

ALIGNMENTS

| RESULT | ID | MANA_STRL1 | STANDARD | PRT | 383 AA. |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|----------|-----|---------|
| AC | P51529 | | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | | |
| DE | MANNA END-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-MANNANASE) (1,4-BETA-D-MANNAN MANNAHYDROLASE). | | | | |
| GN | MANA. | | | | |
| OS | Streptomyces lividans. | | | | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | | | |
| CC | Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces. | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 36-42. | | | | |
| RC | STRAIN-66 / 1326; | | | | |
| RX | MEDLINE; 93207541. | | | | |
| RA | Arcand N., Kluempfel D., Paradis F.W., Morosoli R., Shareck F.; | | | | |
| RT | "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence of the manA gene and characterization of the enzyme."; | | | | |
| RL | Biochem. J. 290:857-863(1993). | | | | |
| RN | [2] | | | | |
| RP | REVISIONS TO C-TERMINUS. | | | | |
| RC | STRAIN-66 / 1326; | | | | |
| RA | Shareck F.; | | | | |
| RL | Submitted (APR-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES CELSIUS AND A PH OF 6.8. | | | | |
| CC | -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS. | | | | |
| CC | -1- SUBUNIT: MONOMER. | | | | |
| CC | -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES). | | | | |
| CC | THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL; M92297; AAA26710.2; -. | | | | |
| DR | INTERPRO; IPR001547; -. | | | | |
| DR | PFAM; PF00150; cellulase; 1. | | | | |
| DR | PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1. | | | | |
| FW | Hydrolase; Glycosidase; Signal. | | | | |
| FT | SIGNAL | | | | |
| FT | CHAIN | | | | |
| FT | DOMAIN | | | | |
| FT | SEQUENCE | | | | |
| SO | MANNA END-1,4-BETA-MANNOSIDASE. | | | | |
| | 36 383 | | | | |
| | 336 340 | | | | |
| | 383 AA; 39681 MW; SDB4B407C64E94C3 CRC64; | | | | |
| | Query Match | | | | |
| | Best Local Similarity | | | | |
| | 28.3%; Score 737.5; DB 1; Length 383; | | | | |
| | 44.4%; Pred. No. 1.8e-43; | | | | |

| Matches | 160; Conservative | 56; Mismatches | 121; Indels | 23; Gaps | | |
|------------|-----------------------------------------------------------------------------|----------------------------------|-------------|----------|--|--|
| QY | 13 LGILLASLLFLFSGSTANAN---- | GFYSGTLLVYANGNPYMRGINGHANYKQDATT | 69 | | | |
| Db | 19 LGILFL-----LAGSAGRAEAAAGSIHNSNGRYGNSAFPMGVNNAHYTPDR-TG | 72 | | | | |
| QY | 70 AIEGIANTGANTVRIYLSDGQGWTKDITVRNLISLADNHLVAPEYDATTG--DSI | 127 | | | | |
| Db | 73 STADIAAKGANTVRYVLSGSGRWTKTSASBSVALIQCKANKVICYLEVHDITTYGKDKA | 132 | | | | |
| QY | 128 ASLNRAVDVWIEKRL--IGKEDTYIINTANFEFGSEGDAMADGYKQAIPLRNAGLNH | 186 | | | | |
| Db | 133 TSLDDQGDYVGVKSSAMPAQEDYVYVNIINGNEFGNTNAAMTADAKSAIGKLRGAGLGH | 192 | | | | |
| QY | 187 TLMVDAAGGQ--FPOSHDYGREVENADPORMFTSIHYEXAGVNASQVRTNIDPVNG | 245 | | | | |
| Db | 199 ALMVDAAPMGQDMGSMGRMNSAVFSADSDRDRTVSEIHNATG-VYDPAAYEDYDLNAFVN | 251 | | | | |
| QY | 246 DLALYIGEEGHHRTNGDVDEATIMTSEQRYGWLAMSKNGKPEWEYLDLSNDWAGNLT | 305 | | | | |
| Db | 252 GPFIYVGEEDGDSHSGNPNEDDAIMATQSLAGVYLGMWSNGGVEYEDMVTNGEFPNEL | 311 | | | | |
| QY | 306 TAMGWTVYGPGLRSTSLSTVFE-TGGGSDGTSPT-----TLVFEESMGWGWS | 356 | | | | |
| Db | 312 TSMGRILYLGSGNIASTATSTATVYGGGGSGTGTAPNGTPIYCVNGASDPDGDGMEWS | 371 | | | | |
| RESULT | | | | | | |
| 2 | | | | | | |
| MANB_CALSA | STANDARD: | PRT: | 1331 AA. | | | |
| ID | MANB_CALSA | P22533: | | | | |
| AC | 01-AUG-1991 (Rel. 19, Created) | | | | | |
| DT | 01-JUL-1993 (Rel. 26, Last sequence update) | | | | | |
| DT | 15-JUL-1999 (Rel. 38, Last annotation update) | | | | | |
| DE | BETA-MANNANASE/ENDOGLUCANASE A PRECURSOR [INCLUDES: MANNAN ENDO-1,4- | | | | | |
| DE | BETA-MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4- | | | | | |
| DE | MANNANASE); ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULOSE)]. | | | | | |
| GN | MANA. | | | | | |
| OS | Caldocellum saccharolyticum (Caldicellulosiraptor saccharolyticus). | | | | | |
| OC | Bacteria: Firmicutes; Bacillus/Clostridium group; | | | | | |
| OC | Thermomonasotacter group; Caldicellulosiraptor. | | | | | |
| RN | [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RX | MEDLINE: 93119139. | | | | | |
| RA | Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.; | | | | | |
| RT | "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a | | | | | |
| RT | multidomain enzyme."; | | | | | |
| RL | Appl. Environ. Microbiol. 58:3864-3867 (1992). | | | | | |
| RN | [2] | | | | | |
| RP | SEQUENCE OF 1-346 FROM N.A. | | | | | |
| RX | MEDLINE: 91247819. | | | | | |
| RA | Luthi E., Jensem N.B., Grayling R.A., Love D.R., Bergquist P.L.; | | | | | |
| RT | "Cloning, sequence analysis, and expression in <i>Escherichia coli</i> of a | | | | | |
| RT | gene coding for a beta-mannanase from the extremely thermophilic | | | | | |
| RL | bacterium 'Caldocellum saccharolyticum'."; | | | | | |
| CC | Appl. Environ. Microbiol. 57:694-700 (1991). | | | | | |
| CC | -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT | | | | | |
| CC | POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH | | | | | |
| CC | MANNANASE AND ENDOGLUCANASE ACTIVITIES | | | | | |
| CC | -1- CATALYTIC ACTIVITY: RANDOM HYDROLISIS OF 1,4-BETA-D-MANNOSIDIC | | | | | |
| CC | LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND | | | | | |
| CC | GALACTOGLUCOMANNANS. | | | | | |
| CC | -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC | | | | | |
| CC | LINKAGES IN CELLULOSE. | | | | | |
| CC | -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES | | | | | |
| CC | CELSIUS. | | | | | |
| CC | -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY | | | | | |
| CC | A (FAMILY 5 OF GLYCOSYL HYDROLASES). | | | | | |
| CC | -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY | | | | | |
| CC | J (FAMILY 44 OF GLYCOSYL HYDROLASES). | | | | | |
| CC | ----- | | | | | |
| CC | THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION | | | | | |
| CC | BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - | | | | | |

[illegible]

| | | | | |
|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-------------------------|---------|
| ID | MANN_BACSM | STANDARD; | PRT: | 513 AA. |
| Dt | PI6699; | | | |
| Dt | 01-AUG-1990 | (Rel. 15, | Created) | |
| Dt | 01-AUG-1991 | (Rel. 15, | Last sequence update) | |
| Dt | 01-FEB-1991 | (Rel. 17, | Last annotation update) | |
| DE | MANNA ENDO-1,4-BETA-MANNOSEDASE A AND B PRECURSOR (EC 3.2.1.78) | | | |
| DD | (BETA-MANNANASE) (ENDO-1,4-MANNANASE). | | | |
| O5 | Bacillus sp. (Strain AM-001). | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/staphylococcus group; Bacillus. [1] | | | |
| RN | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE. | | | |
| RP | MEDLINE: 90146329. | | | |
| RX | Akino T.; Kato C.; Horikoshi K.; | | | |
| RA | "Two Bacillus beta-mannanases having different COOH termini are produced in Escherichia coli carrying pMAH5." | | | |
| RL | Appl. Environ. Microbiol. 55:3178-3183(1989). | | | |
| Cc | -I- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS. | | | |
| Cc | -I- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES. | | | |
| Cc | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outbreak station - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isdb.ch/announce/or send an email to license@isdb.ch). ----- | | | |
| CC | EMBL: M31797; AAA22586.1; -. PIR: A37219; A37219. | | | |
| DR | INTERPRO: IPR000805; .- | | | |
| DR | PRINTS: PR00739; GLEHYDLRSLSEZ6. | | | |
| KW | Hydrolase; Glycosidase; Signal. | | | |
| FT | SIGNAL 1 26 | | | |
| FT | CHAIN 27 513 | | | |
| FT | CHAIN 27 365 MANNA ENDO-1,4-BETA-MANNOSIDASE A. | | | |
| SO | SEQUENCE 513 AA; 58430 MW; 8BD105F622CDB5A8 CRC64; | | | |
| | | | | |
| Query Match 16.4% Score 427; DB 1; Length 513; | | | | |
| Best Local Similarity 26.9%, Pred.No. 4,3e-22; | | | | |
| Matches 149; Conservative 76; Mismatches 215; Indels 114; Gaps 23 | | | | |
| OY | 5 FKIKSIITLSLLASILEFGSGTSAMNSGYYVSGTTLYDGNGPFVV----- | 53 | | |
| Db | 4 YKRVAFVAFIHFESVLPTISMSTEAN-----GALSNRNAGQTYTMNYSWLANLPN | 55 | | |
| OY | 54 ---KGINHHGH--AWMYDAQTAIEGIANTGANTVRIVLS--DGGOW-----TKDDIHVTVN- | 102 | | |
| Db | 56 KSNNKVSVGSHEGFSDSTLAWKIOCARLETGKMPGGITSCDYKNMQPRLVADQISYGNCQ | 115 | | |
| OY | 103 -LIISAEDNHVIAY-----PEVDHATGYDSIASLRANDY-----MTEKSALI 145 | | | |
| Db | 116 ELTNFNWGOGTLTVISHVMNPFGFHGEENKTILLPTSOFNLNTNRHTTGRRRWKDMLDKWA | 175 | | |
| OY | 146 GKED-----TWINIANIEMFGSWGEGDW--ADGTKOAIPLRLNAGLN-----H 186 | | | |
| Db | 176 DGDLDQLONNGVYLVLRPLRHENGEM---FWGAEEGINODQRANKYIAAMDAMIOTFYTH | 232 | | |
| OY | 187 TLMDAAGMGFPQSIIHDYGREVFNADPOKMTFSIMHETLAGNASQVRTNIDRYLAND | 246 | | |
| Db | 233 EKRLNNLIMVYSBPDIYRDH---VTSYPGANVYDIVALDSY-HPDHSITDQYNRMIALD | 288 | | |
| OY | 247 LALLVGEGEHRT--NGDVDEATIMSSEQ---RGCGKLAMS--WR--GNGPMEYEYIDLSEN | 298 | | |
| Db | 289 KPFAFAEIGPPESMASSFEYSNIQAIRKYDPRIYFLVALNDKWSPHNRRGM--DFLN 345 | | | |

[illegible]

[illegible]

| ID | GEN5_BACAG | STANDARD; | PRT; | 400 AA. |
|----|----------------------------------------------------------------------------------------|-----------------------------------|------|---------|
| AC | 085465; | | | |
| AD | 15-JUL-1999 | (Rel. 38, Created) | | |
| DT | 15-JUL-1999 | (Rel. 38, Last sequence update) | | |
| DT | 30-MAY-2000 | (Rel. 39, Last annotation update) | | |
| DE | ENOGLOUCANASE 5A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (ALKALINE CELLULOSE). | | | |
| DE | CEL5A. | | | |
| OS | Bacillus agaradhaerens (Bacillus agaradhaerens). | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| OC | Bacillus/staphylococcus group; Bacillus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-DSM 8721; | | | |
| RA | Bjoernvad M.E.; | | | |
| RL | Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329. | | | |
| RC | STRAIN-AC13 / NCIMB 40482; | | | |
| RX | MEDLINE: 98153671. | | | |
| RA | Davies G.J., Dauter M., Brzozowski A.M., Bjoernvad M.E., | | | |
| RA | Andersen K.V., Schuelein M.; | | | |
| RT | "Structure of the Bacillus agaradhaerens family 5 endoglucanase at 1.6 | | | |
| RT | A and its cellobiose complex at 2.0-A resolution." | | | |
| RL | Biochemistry 37:1926-1932(1998). | | | |
| RN | [3] | | | |
| RP | X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329. | | | |
| RC | STRAIN-AC13 / NCIMB 40482; | | | |
| RX | MEDLINE: 98384138. | | | |
| RA | Davies G.J., Mackenzie L.F., Varrot A., Dauter M., Brzozowski A.M., | | | |
| RA | Schuelein M., Withers S.G.; | | | |
| RT | "Snapshots along an enzymatic reaction coordinate: analysis of a | | | |
| RT | retaining beta-glycoside hydrolase." | | | |
| RL | Biochemistry 37:11707-11713(1998). | | | |
| CC | -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE. | | | |

CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>).
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF067428; AAC19169.1; -
DR PDB: 1A3H; 16-MAR-99.
DR PDB: 2A3H; 16-MAR-99.
DR PDB: 3A3H; 16-MAR-99.
DR INTERPRO: IPR001547; -
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT SITE 165 165
FT ACT SITE 254 254 NUCLEOPHILE.
SQ SEQUENCE 400 AA; 44702 MW; 399C66FBC36FEF CRC64;

```

Query Match 8.8% Score 229.5 DB 1 Length 400;
Beet Local Similarity 23.4% Pred. No. 9.5e-09;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

QY 6 KRIFFSTILLASSILFEVSGSTANANS-----GFYVSGTTLTDANGNPFVMKGI-NH 58
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 2 KATTEIVLVLTALFEISIGNTTADNDNSVHEHGOLSTISNGELVNERGEQVQLGMSGH 61
QY 59 GMAVT-----KDAATTALEGLANTANTVRYTL--SDSGQWTKDDI-HIVRL 103
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 62 GLOWGQFVNYSMKRLR-----DWGINFRAMYTSSGQYIDDPVKKEKKA 111
QY 104 ISLAEDNHLVVAPEVHADGQDYSIASINRAVDYWIEMRSALIGKEEDYIINIAEWFGSM 163
DB 112 VERAIDLDIYIYIIDHILSDNDPNITKEKAKPFDDM-SELGQDPNVYIELANEPNGS- 169
QY 164 EGDMAADGK---QAIPLRNAGLHTLMDVDAAGNGQFPQSIHDYGREVFNADPQRTM 219
DB 170 -DVTGNOQIKRPAEEVPIIRNNDPNNIIIV--GRTGWSQDVH-HAADNQLADP--NYM 222
QY 220 FSTHMYEYVGNAQRYRTIDRYLNDLALVIGEGHRTNDP---VDEATI-MGISEO 274
DB 223 YAFHF--YAGTIGQNRDQVDYALDQGAIEVSEWTSATGDGCVFLDEAOYWDIFMDE 280
QY 275 RGVGWLAWS-----WKGNGPEWEYLDLSNDMAGNNLTJWAGNTIVNGEYGLRETS 323
DB 281 RNLSTANWELFHNDESSALMPGAND-----TGCTEALSLSPSGTFVRE--KIRESA 330
QY 324 RUSTYETGGGSDGSGPTTLTYDEFGSMQWTSGL-----SGGPAVATWS 369
DB 331 SIIPSDPTPPSPGEPDPTPPSDPGETPAMPDQIITNEIVYHNGLWQAKMT 384

RESULT 6
GUN1_BAC34
ID GUN1_BAC34 STANDARD; PRT; 488 AA.
AC P06566;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE).
GN CELA.
OS Bacillus sp. (strain N-4).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE: 87056924.

```

RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RL *Bacillus* sp. strain N-4 and their strong homology.";
 CC J. Bacteriol. 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC The SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14781; AAA2301.1; -
 CC PIR: A25156; A25156.
 CC INTERPRO: IPR001547; -
 CC PFM: PF00150; cellulase: 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC KMW Cellulose degradation; Hydrolase; Glycosidase.
 CC FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
 CC FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
 CC SEQUENCE 488 AA: 54264 MW: 972488453028385 CRC64;

Query Match 8.7%; Score 226.5; DB 1; Length 488;
 Best Local Similarity 23.1%; Pred. No. 2e-08;
 Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;

QY 6 KRIFFITSLIASSILFVSTGTSTAN-----ANSGFVSGTLLYDANGNPFVARGI-NH 58
 DB 2 KKITITFIYFLTLA--LLEFGNSTANNGSVNGQLSIOGQLVNEHGDVPLQKMSH 59
 QY 59 GHAWYKDOATTAEIGANTGANTVRIYLSGGQWTKD-----IHT----- 99
 DB 60 GLQWYG-----QFVNDSTIKMLDDMGITVFRAMYTSSGGYIEDPS 101
 QY 100 ----VRNLISLAEDNHLVAVPEVDATGYDSIASLNRAVDYIMERSALIGKEDYIINI 155
 DB 102 VKEKVEKEAVEAIDIGIYIIDMHILSDPNYIKKEAEFEDEM-SALXGDPVNYIYI 160
 QY 156 ANEWFG--SWEGDAWADGY-KOALPRLNAGLNHTLMDAAGWQFPOSIDYGREVN 211
 DB 161 ANEPGNHNVYR--DSHIKRYAEVPIVIRANDPNNIYIVGTATWS--QDVHEADNQD 215
 QY 212 ADPOQNTMESIMYERYAGNASOVRTNDRVYNODLALVIGFGRHTNGD-----VDEN 267
 DB 216 -DP--NVMYAFHF--YAGTHGOQLRQVDYALSRGAALIFVSEWGTSAATGDGVFLDE 270
 QY 268 I-MSYSEQRGVGLAWS-----WKNGPEWEYIDLSDNMGNNLTANGNTIYNG 315
 DB 271 VMDIDMDEKRLSWANWSLTHKDESSAALMPGASP-----TGGWTAELSSGAFVRE- 322
 QY 316 PYGLRETSRLSTVFTGGSD---GGTSPTLLYDFEGSMQWTSGLS-----SGGPPA 364
 DB 323 --KIREASITPPSDPTPPSDPPGEPDPFPSPD-PGEYRAMPNQIYTNIEIYHNGQLNQ 379
 QY 365 VTEM 369
 DB 380 AKWMT 384

RESULT 7
 GUN2_BACSA
 ID GUN2_BACSA STANDARD; PRT; 409 AA.
 AC P06565;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE).

GN CELL.
 OS *Bacillus* sp. (strain N-4).
 CC Bacteria; Firmicutes; *Bacillus*/Clostridium group;
 CC *Bacillus*/Staphylococcus group; *Bacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87056924.
 RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
 RT "Nucleotide sequences of two cellulase genes from alkalophilic
 RL *Bacillus* sp. strain N-4 and their strong homology.";
 CC J. Bacteriol. 168:479-485(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC The SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14729; AAA2299.1; -
 CC PIR: B25156; B25156.
 CC INTERPRO: IPR001547; -
 CC PFM: PF00150; cellulase: 1.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC KMW Cellulose degradation; Hydrolase; Glycosidase.
 CC FT ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
 CC FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
 CC SEQUENCE 409 AA: 45690 MW: 988866066B4DADF CRC64;

Query Match 8.7%; Score 226; DB 1; Length 409;
 Best Local Similarity 23.2%; Pred. No. 1.7e-08;
 Matches 104; Conservative 62; Mismatches 147; Indels 136; Gaps 2;

QY 6 KRIFFITSLIASSILFVSTGTSTA-----NANSGFVSGTLLYDANGNPFVARGI-N 57
 DB 2 KKITITFIYFLTLA-LFIIGNTAADYSVYEEHGLSISNGELVNDGCEYQKMS 60
 QY 58 GHAWYKDOATTAEIGANTGANTVRIYLSGGQWTKD-----IHT----- 99
 DB 61 HELQWYG-----QFVYESMKMLRDDMGITVFRAMYTSSGGYIEDP 102
 QY 100 ----VRNLISLAEDNHLVAVPEVDATGYDSIASLNRAVDYIMERSALIGKEDYIINI 154
 DB 103 SVKEKVEKEAVEAIDIGIYIIDMHILSDPNYIKKEAKDEFDEM-SELYGDYENYIYE 161
 QY 155 IANEWFGS--IWEGD--AWADGYKOALPRLNAGLNHTLMDAAGWQFPOSIDYGREV 209
 DB 162 IANEPNGSDVTNDQIKPAE--EVIPIYRNDENNIITV--GTGMSQDVH-HAADN 214
 QY 210 ENADPOQNTMESIMYERYAGNASOVRTNDRVYNODLALVIGFGRHTNGD-----VDE 265
 DB 215 QLTDP--NVMYAFHF--YAGTHGOQLRQVDYALOGAALIFVSEWGTSEATGDGVFLDE 270
 QY 266 ATI-MSYSEQRGVGLAWS-----WKNGPEWEYIDLSDNMGNNLTANGNTIY 313
 DB 271 AQVWIDFMDERLISWANWSLTHKDESSAALMPGASP-----TGGWTAELSSGFTV- 322
 QY 314 NCPYGLRETSRLSTVFTGG-----SDGCT-----SPTLLYD 345
 DB 323 -----REKIRSATTPSDPTPPSDPPGEPDPFPSPDPPGDPYAMDENTIT 376
 QY 346 FE-----GSM--OGWTSGLSG---GFW 363
 DB 377 DEIVYHNGQLMOAKWMTQNOEPGDYGPW 405

RESULT 8

GUNW_ERMCA STANDARD: PRT: 505 AA.

ID GUNW_ERMCA 047096: 01-NOV-1997 (Rel. 35, Created)

AC 047096: 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE ENDOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)

DE (CELLULOSE V).

GN CELV.

OS Erynia carotovora.

OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;

OC Pectobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SCRI193;

RX MEDLINE: 94067016.

RA Cooper V.J.C., Salmond G.P.C.;

RT "Molecular analysis of the major cellulase (Celv) of Erynia carotovora: evidence for an evolutionary 'mix-and-match' of enzyme domains."

RT Mol. Genet. 241:341-350(1993).

CC -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES CELSIUS.

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).

CC -----

CC EMBL: X76000; CAA53592.1; -

DR HSSP: Q06851; INBC.

DR INTERPRO: IPR001547; -

DR INTERPRO: IPR001956; -

DR PFAM: PF00942; CBD_3; 1.

DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

KW SIGNAL

FT CHAIN 1 31 POTENTIAL.

FT DOMAIN 32 505 ENDOGLUCANASE V.

FT DOMAIN 335 352 CATALYTIC.

FT DOMAIN 353 505 LINKER.

FT ACT_SITE 168 168 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 256 256 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

FT SEQUENCE 505 AA; 54900 MW; DBEA9337B4D2623 CRC64;

Query Match 8.5%; Score 220.5; DB 1; Length 505;

Best Local Similarity 21.8%; Pred. No. 5.3e-08;

Matches 116; Conservative 69; Mismatches 209; Indels 137; Gaps 21;

2 NNGFKKFTSLTSLILASLIFVSGTSTANANGS-FYVSGTTLTDANGNPFVAKGI-NHG 59

6 NOYRKRLITGVTVVLCMSISFSLSTPETHGOLSIENGRILVDEGRKVOQLRGISSHG 65

60 HAWKIDQATTAIEGIANTGANTVAVIISDGGQMTKDD----- 96

66 LQWFGD-----YVKKDSMKWLRDMDGIVNFRVANYTADGYISNPSTL 107

97 IHTVRLNISLAEDNHLVAVPEVHATGYDSIASLNRAVDVMIEMRSALLIKEDVYIINIA 156

108 ANKYEVAANAQSLGVITITIDMHLISNDPNTYKAQKTFPAEL-AQLYSSSPVAYIEIA 166

157 NEMFG--SWEGD--AWADGYKQALPRLNAGLNTLAVYDAAGWGFPOSIDHGREVENA 212

DB 167 NEPGVYTWNGQIPYA---LEVTDTIRSKDPPMLIV---GISTWQDHLHD---AADNQ 217

QY 213 DPORTMESIMPEYAGGNASQVNTINDRYLNDLALVIGFGRHTNGD-----VDEAT 267

DB 218 LPDNTMYALHF--YAGTHGQFLDRIDYAOISGALFVSEWGTSDASNGPFLPESQT 275

QY 268 IMTSSEORGVWLMASKNGNPEWEYL-----DLSMDAGNLTLMGVTIYNGPYGLRETS 323

DB 276 WIDFLNNGVSMVNMSTLTKDSEASALAPGASKGSGWTEONLSTSGKVV-----REQI 328

QY 324 RLSVFTGSGSDGTSPTLLDYDFEGSMGWTSSLSGGPMNAVTEMSSKSHSLADIOLS 383

DB 329 R-----AGANNGGDPPTTPTPTPTPTPTPGCTTG-----DYYL- 359

QY 384 SNSQHYLVHONTSLQONSRIQATVKHANGSVG--NGMTARLYVTGHCYTWSSGFV 440

DB 360 ---QYRVNDNPPS---DDAIRMAVINIKTGSTPIKLSDLQVRYFHD----- 400

QY 441 PINGSGTSLDLSNTQNSQVREIGVQFSAD-----SSGQTSI 482

DB 401 -DGRGANLFEVDWANNPNNIVTGTTPAASTRKANRYLVYVSSGAGSL 449

RESULT 9

GUNW_ERMCA STANDARD: PRT: 444 AA.

ID GUNW_ERMCA 059394: 01-NOV-1997 (Rel. 35, Created)

AC 059394: 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE ENDOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N)

DE (CELLULOSE N).

GN CELN.

OS Erynia carotovora.

OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;

OC Pectobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATROSEPTICA FCBR C18;

RX MEDLINE: 9829944.

RA Olsen O., Thomsen K.K., Weber J., Dues J.O., Svendsen I., Wegener C., von Wettstein D.;

RT "Transplanting two unique beta-glucanase catalytic activities into one multienzyme, which forms glucose."

RT Biotechnology 14:71-76(1996).

RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).

CC -----

CC EMBL: I39788; AAC37033.1; -

DR INTERPRO: IPR001547; -

DR INTERPRO: IPR001956; -

DR PFAM: PF00942; CBD_3; 1.

DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

KW SIGNAL

FT CHAIN 1 31 POTENTIAL.

FT DOMAIN 32 444 ENDOGLUCANASE N.

FT DOMAIN 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

FT SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 8.4%; Score 219.5; DB 1; Length 444;
 Best Local Similarity 21.7%; Pred. No. 5.2e-08;
 Matches 108; Conservative 71; Mismatches 193; Indels 125; Gaps 21;

2 NNGFKIFSTLTLASSILFVSGSTANANSQ-FYVSGTLYDANGNPFVARGI-NHG 59
 6 NOYAKLLGVTTVLGMSLSFSAISATPVEFHGOLSTINGLVDPQKRVGLRGVSSHG 65
 60 HAWYKQATTAIEGIANGTAVRIYLSGGQWTRD----- 96
 66 LQWFGD-----YVKNDSMKWLRDWMGINVERVAMYTADGYISNPGL 107
 97 IHTVNLISLAEDNHLVAVPEVHDATGYDSTASLRNADYIEMKSAIIGKEDYIINIA 156
 108 ANKVEAAVAAAGSLGVYIIIDWHILSDNDPNYKQAKIFFEM-AGYIGSSPNYIYELA 166
 157 NEMFG--SMEGD--AMADGYKQAIPLRNAGLHNTLVDAAGWGPPOSIDHGYREVENA 212
 167 NEPNQGVATWNGQIRPYA---LEVDTIRSKDPDNLIIY---GTGWSQDIDH---AADNQ 217
 213 DQQRNTMSIHYEYAGNASQVRTNIDRVLNQDLALYIGFGRHRTNGD-----VDEAT 267
 218 LPDPRTLYALHF--YAGTHGQFLRDIRDYAQSRAAIFVSEMGTSDASGNGPFLPESQT 275
 268 IMYSIQGVGWTANSMKNGNPEWEYL-----DLSNDAGNNTIANGNTIYNGPYGLRETS 323
 276 WIDFNNRGVSWVWNSLSDKSEASNAALPAGASKSGWTEQNSTGSKFY-----REQI 328
 324 RUSTFTGGSGSGTSPPTLYDFEGSMQWGTSSLSGPMWATEWSSKSHSLAKDIQLS 383
 329 R-----AGANTGGGDTPTT-----PTTPEPTNP-NGTTGDVYL- 362
 384 SNSQHLVIONTSIQQSRIGATYKHAMWGSVG---NGMTARLYKTHGHTYSGSEY 440
 383 ---QYRVNDNPS---DDAIRMAFNINKTGSTPIKLSLDQVRYFHD----- 403
 441 PINGSGTTLSDLSNV 457
 404 -DGRGANLFVDMANV 418

RESULT 10
 GUNW_ERMCA STANDARD; PRT; 504 AA.
 AC 059395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ENDOGLUCANASE VI PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE VI)
 DE (CELLULOSE VI).
 GN CELV1.
 OS Eryinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SCC3193;
 RX MEDLINE; 95231512.
 RA Mae A., Helsinki R., Palva E.T.;
 "Structure and regulation of the Eryinia carotovora subspecies
 carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
 phytopathogenicity";
 Mol. Genet. 247:17-26(1995).
 CC -I- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -I- SUBCELLULAR LOCATION: SECRETED.
 CC -I- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in the
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb.ch/announce/>
 or send an email to license@isb-eb.ch).

DR EMBL; X79241; CAA55823.1; -
 DR HSSP; 006851; INBC
 DR INTERPRO; IPR001547; -
 DR INTERPRO; IPR001956; -
 DR PFAM; PF00942; CBD_3; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosylase; signal.
 FT SIGNAL 1 31
 FT CHAIN 32 504
 FT DOMAIN 32 334
 FT DOMAIN 335 352
 FT DOMAIN 353 504
 FT ACT_SITE 168 168
 FT ACT_SITE 256 256
 FT ACT_SITE 256 256
 FT ACT_SITE 256 256
 SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 7.9%; Score 207; DB 1; Length 504;
 Best Local Similarity 22.3%; Pred. No. 4.4e-07;
 Matches 110; Conservative 63; Mismatches 191; Indels 130; Gaps 22;

6 KRIEFTLSLTLASSILFVSGSTANANSQ-FYVSGTLYDANGNPFVARGI-NHGAMY 63
 10 RULTLGVTTLMGMSLSFSAISATPVEFHGOLSTINGLVDPQKRVGLRGVSSNGLV 69
 64 KDQATTAIEGIANGTAVRIYLSGGQWTKD--IHTVNLISLAEDNHLVAVPEV--- 118
 70 GD-----YVKNDSMKWLRDWMGINVERVAMYTEAE-NGYIANPSLANK 110
 119 -----HDAITYDSTASLRNADYIEMKSAIIGKEDYIINIANEM 159
 111 VKEAVAAAGSLGVYIIIDWHILSDNDPNYKQAKAIFFAEM-AGYIGSSPNYIYELANP 169
 160 FGS--WEGD--AMADGYKQAIPLRNAGLHNTLVDAAGWGPPOSIDHGYREVENADPO 215
 170 NSGYATWNGQIRPYA---LEVDTIRSKDPDNLIIYSGSGWS---QDIDH---AADNQLPD 220
 216 RNTMFSIMHYEYAGNASQVRTNIDRVLNQDLALYIGFGRHRTNGD-----VDEATIMS 270
 221 PNTLYALHF--YAGTHGQFLRDIRDYAQSRAAIFVSEMGTSDASGNGPFLPESQWTID 278
 271 YEDQGVGWTANSMKNGNPEWEYL-----DLSNDAGNNTIANGNTIYNGPYGLRETSRLS 326
 279 FLNNGGISWVWNSLSDKSETSAALVAGASKSGWTEQNSTGSKFY-----REQIR-- 329
 327 TVEFTGGSDGSGTSPPTLYDFEGSMQWGTSSLSGPMWATEWSSKSHSLAKDIQLSNS 386
 330 ---AGAGLSGGDTPTTPEPTNPNGTIG-----DYL----- 359
 387 OHYLVIONTSIQQSRIGATYKHAMWGSVG---NGMTARLYKTHGHTYSGSEFVPIN 443
 360 -QYRVNDNPS---DDAIRMAFNINKTGSTPIKLSLDQVRYFHD-----D 401
 444 GSSGTTLSLDLSNV 457
 402 GRPGANLFVDMANV 415

RESULT 11
 GUNL_BACSTU STANDARD; PRT; 499 AA.
 AC P07983;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE).


```

GN  BGIC OR GID.
OS  Bacillus subtilis.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/staphylococcus group; Bacillus.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-DLG:
RX  MEDLINE: 87194581.
RA  Robson L.M., Chambliss G.H.;
RT  "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
RL  J. Bacteriol. 169:2017-2023(1987).
CC  -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC  LINKAGES IN CELLULOSE.
CC  -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC  HYDROLASES).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M16185; AAA2496.1; ALT_INIT.
DR  PIR: A26874; A26874.
DR  HSSP: Q06851; INBC.
DR  INTERPRO: IPR001547; -.
DR  INTERPRO: IPR001956; -.
DR  PFM: PF00942; CBD_3; 1.
DR  PFM: PF00150; cellulase; 1.
DR  PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW  Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT  SIGNAL 1 29
FT  CHAIN 30 499
FT  ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT  ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT  DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT  SEQUENCE 499 AA; 55187 MW; 339D04EE95A63E1 CRC64;

Query Match 7.2%; Score 188.5; DB 1; Length 499;
Best Local Similarity 21.1%; Pred. No. 8e-06;
Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17;

OY 6 KRIFITITSLASSILF-----VSGISTANANSQ-FYVSGTTLTDANGNPFVWNG 55
DB 2 KRISIFITCLIAVLTMGILPSPASAGTTPVAKNGQLSIRGTQLVNRDCKAVQLNG 61
OY 56 I-NHGSAWKQDQATVAIEGANTGANTVATVLDGQGTOKD-----IHT----- 99
DB 62 ISSHLOWTGD-----FVKDSIAKLRDMDGIVTFAAAYTADGVI 103
OY 100 -----VRNLSIAEDNHLVAPEVDHATGYDSIASINRAVDYWIEMRSALIGREDTV 151
DB 104 DNPVYKNKKEAWEAKGIVIIIDMHLNDGNPNQNEKKEFEKEM-SLYGTPTNV 162
OY 152 IININEMNG--SWEGD--AMADGYKQATPRLRNAGINTLAVDAAGWQFQSHIDYR 207
DB 163 IYEINNEPNDVNMKRDIPYAE---EVSIVRKNDPDIIIV---GGETWSDQVND--- 213
OY 208 EVFNADPQ---NNTFSSIMYEYAGGNASQVRTNIDRVNODLAVIGEGRRHTNGD-- 262
DB 214 ---AADDOUKDANVMYALHF--YAGTHGSLDKXKANYALSKAPFIVYEMGSDASGNG 268
OY 263 ---VDEA-TIMSTSEORGWGLAMSKGNGPEWEYL---DLSNDNAGNNLTMAGNTIYNG 315
DB 269 VFLDSSRETLNIDSKNISWVMNLSDKOESSALKPKGASKGTGWPVLTDLTASGTFVREN 328
OY 316 PYGLRETSR 324
DB 329 INGTDRSTK 337

```

```

RESULT 12
GN2_BACSU STANDARD; PRT; 499 AA.
ID GUN2_BACSU
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMC-ASE) (CELLULOSE).
GN BGIC OR GID OR BGIS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PAP15;
RX MEDLINE: 87066783.
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RT Moraneill F., Seligy V.;
RL "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CK-2;
RX MEDLINE: 95225656.
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RL subtilis CK-2.";
RL Antoine Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Rose M., Ertlan K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 30-45.
RC STRAIN-CK-2;
RX MEDLINE: 95225655.
RA Aa K., Flengsrud R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RT compost soil.";
RL Antoine Van Leeuwenhoek 66:319-326(1994).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z29076; CAAB82317.1; -.
DR EMBL: X04689; CAAB82317.1; -.
DR EMBL: X67044; CAAB7429.1; -.
DR EMBL: Z73234; CAAB7429.1; ALT_INIT.
DR EMBL: Z99113; CAB13696.1; ALT_INIT.
DR PIR: A26114; A26114.
DR HSSP: Q06851; INBC.
DR SUBTILIST; BG10437; BGIC.
DR INTERPRO: IPR001547; -.
DR INTERPRO: IPR001956; -.
DR PFM: PF00942; CBD_3; 1.
DR PFM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.

```

FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 FT CONFLICT 283 283 S -> N (IN REF. 2).
 SQ SEQUENCE 499 AA; 55287 MW; 8F735FE71B3EAE2 CRC64;

Query Match 6.9%; Score 181; DB 1; Length 499;
 Best Local Similarity 19.3%; Pred. No. 2.6e-05;
 Matches 100; Conservative 102; Mismatches 188; Indels 128; Gaps 26;

QY 6 KKFTSLSLASSILF-----VSGTANANS-GFYSGTLLYDANGPEFMARG 55
 DB 2 KRISITITLTLTLTMGMASIPASACTKTPVANKNGLSIKGTLVNRDGAVALKG 61
 QY 56 I-NHGANWKDQAT-TAIEGIANGTANTV--RIVSDGGQWTRDDI-HTRNLTSLAED 109
 DB 62 ISSHGLQWGEYVNDLSKFLRDDMGITVRAMYTADGGYIDNPVKNVKEAVEAKE 121
 QY 110 NMLVAVPEVDATGYDSIASLNAVVDYIEMRSALIGKETVYIINTANEMFG--SWEGD- 166
 DB 122 LGIYIIVDMLNDGNPNQKKEKFEKEM--SSLGNTPTNVIYELANENGDNMRDI 180
 QY 167 -ANADGYKQAIPLRNAGLNHTLVDAAGQGFPOSIDHYGREVENADPQ---RTMFSI 222
 DB 181 KPYAE---EYISYIKRNDPNIITV---GTGTSQDVND-----AADDLKANVYAL 228
 QY 223 HMYEYAGNASQVRTNIDRYLNODLALVIGFGRHRTNGD---VDEA-TIMSYSEQGV 277
 DB 229 HF--YAGTHQOFLRDKANVALSKGAPLEVTWEGTSDASGNGVFLDQSRMLKYLDSKI 286
 QY 278 GWLANSKNGPEWEYL---DLSNDAGNNLTAANGTYNGRYGLAETSR----- 324
 DB 287 SWNMNLSDKOESSALKPGASKGTGRLSDLSAGTFVENILGTQSTKDIPEPESKD 346
 QY 325 -----LSTVFTGGSDGTSPTTLYDFEGSMOGMTGSSLSGPNVATEMSSGSHLK 377
 DB 347 KPYENGISVQYRAG-----DGSN-----NSQIR 371
 QY 378 ADIQLSSN-----SQHYLVIONTSLQONSRIQATVKRANMGVNGMTARLYV- 426
 DB 372 POLQIKNGMTYVDLKDVTARYWKAKNKG--QNF-----DQDYAQICGNVTHKFTV 422
 QY 427 --KTGHGYTY-----SGSFYPIGSSGTTLSLDSN 456
 DB 423 LHRKQGAADTYLELGFKNGLIAP--GASTGNIQLRLIN 458

RESULT 13
 GUN5_THRFU STANDARD; PRT; 466 AA.
 ID GUN5_THRFU
 AC 001786;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4) (CELLULOSE E-5) (CELLULOSE E5).
 GN CELE.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TX;
 RX MEDLINE; 91258320.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT "DNA sequences of three beta-1,4-endoglucanase genes from
 RT Thermomonospora fusca".
 RL J. Bacteriol. 173:3397-3407(1991).
 RP REVISIONS.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE OF 137-142 AND 157-166.
 RC STRAIN-TX;
 RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
 RT "Activity studies of eight purified cellulases: specificity,
 RT synergism, and binding domain effects.";
 RL Biotechnol. Bioeng. 42:1002-1013(1993).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; L01577; AAC09379.1; -
 DR PIR; C42360; C42360.
 DR HSSP; P07986; 1EXH.
 DR INTERPRO; IPR001547; -
 DR INTERPRO; IPR001919; -
 DR PFAM; PF00553; CBD_2; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT CHAIN 1 36
 FT ACT_SITE 37 466
 FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 466 AA; 49800 MW; 1CF0ADBEF2DEF82E CRC64;

Query Match 6.9%; Score 180; DB 1; Length 466;
 Best Local Similarity 25.0%; Pred. No. 2.8e-05;
 Matches 82; Conservative 49; Mismatches 129; Indels 68; Gaps 19;

QY 26 GSTANANS-GFYSGTLLYDANGPEFMARGIN-HGHWYKQDQAT-TAIEGIA-NTGANT 81
 DB 160 GTGTPVERYKRVYCGTQCDDEHGNPYQLKGMSTHGIQWFDHCLTDSLDALAYDKADY 219
 QY 82 VRI---VLSDGGQWT---KDDIHTVKNLISLAEDNHLVAVPEVDATGYDSIASINRAV 134
 DB 220 IRLSMYIQEDGYEFINPRGFDRH---QLIDMATAGLVYIVDWHILTPDPHYNIDRAK 276
 QY 135 DYIEMRSALIGKEDYIINIANEMFGSMEGDAMAD--GI-KQAIPLRNAGLNHTLMDV 191
 DB 277 TFFAEIAORHASKTN-VLAEIANE---PNGVSMASIKSYAEVYIPRIORDPDSVYIYG 331
 QY 192 AAGMGQFPOSIDHYGREVEFAAD--QRTMFSIMTYEYAGNASQVRYTNDVYLNODLAL 249
 DB 332 TRKMSLSGVS-BESGFAELIANPVYASNTIMTAHFY-----AASHRDYLNALREASEL 364
 QY 250 ---VIGFGRHRTNGD-----VDEATISYSEORGVGLANSWK----- 285
 DB 385 FPFVTEVEFGRETYTGOANDFOVADR--YIDLAEKRIGIKTNNYSDDFSGAVFOPGTC 442
 QY 286 -GNGPEWEYDLSNDWAGNLTAWGNTI 312
 DB 443 ASGGP-----WSSGLKASGQWV 460

RESULT 14
 GUN3_BACSTU STANDARD; PRT; 499 AA.
 ID GUN3_BACSTU
 AC P23549;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULOSE) (CMCase) (CELLULOSE)
 GN BGLC.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BSE616;
 RA MEDLINE; 9129280.
 RX Park S.H., Kim H.K., Pack M.Y.;
 RT "Characterization and structure of the cellulase gene of Bacillus
 subtilis BSE616.";
 RL Agric. Biol. Chem. 55:441-448(1991).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D01057; BAA00859.1; -.
 DR PIR; JN0111; JN0111.
 DR HSSP; 006851; INBC.
 DR INTERPRO; IPR001547; -.
 DR INTERPRO; IPR001956; -.
 DR PRAM; PF00942; CBD_3; 1.
 DR PRAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR CELLULOSE degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499 ENDOGLUCANASE.
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 SO SEQUENCE 499 AA; 55169 MW; 2E821E3D8BAC04 CRC64;

Query Match 6.68; Score 171; DB 1; Length 499;
 Best Local Similarity 19.08; Pred. No. 0.00013;
 Matches 98; Conservative 104; Mismatches 189; Indels 126; Gaps 25;

QY 6 KRIEITLILASSILF-----VSGSTANANG-PPVSGTTLTYDANGPEVWAG 55
 DB 2 KRISITITCLITLIMGMLASPAASAGTKITPAKNGQSLSTGQLVNNDGKAVQLKG 61
 QY 56 I-NHGHMYKQAT-TALEIGANTGANTV--RYLSDGGGWTKDDI-HYRNLISLAED 109
 DB 62 ISSHGLQWYGEYVKNKSLKMLRDMGIVFRAMYTADGGIIDNPVKNKKKEAVEAKE 121
 QY 110 NMLVAVEVHATGYDTSASINRAVDYIEMRSALIGEDYIINIANEMWG--SWED- 166
 DB 122 LGIVYIIDLHMLNDGNPNQNEKKEFEKEM-SSLYGNTPVYIEIAEPNGVDNWKMDI 180
 QY 167 -AMADGYKQAIPLRLNAGLINTLAWDAGMGQFPOSIDYREVENADPQ--RNTMESI 222
 DB 181 KPYAE---EYISVIRKNDPMIITV---GTGTMSQDVND-----AADDQLKANVMDAL 228
 QY 223 HMEYAGNAGSOVNTNIDRVINQDLALVIGEGHRTNGD---VDEA-TIMSYSEOGCV 277
 DB 229 HF-YAGHGOFLDKKANYALSKGAPIVTEWGTSDASGNGGVFDQREMLKYDLSKTI 286
 QY 278 GWLAWSTWNGNPEWEYL---DLSNDWAGNNTLNGNTIVNGPYGLRTSIR----- 324
 DB 287 SWVMNMLSDKQESSALKPGASKTGGMRLSDLSASGITVERENILGTOKSTADIPETPAKD 346
 QY 325 -----LSTVFTGGSGDGTSPITLYDFEGSMQGTGSSLSGPFWAVTEWSSKSHSLK 377

DB 347 KPTQENGISVOYAG-----DGSN-----NSNOIR 371
 QY 378 ADIQLSNSQHYLHVQIQLNSIQ-----QNSRIQATVKHANGSVNGMTARLYV-- 426
 DB 372 PQQLINNG-----NTVYDLKDYATARYWYAKKNGQVDDYADLGQGVNTYFVYL 423
 QY 427 -KTGHGYTWY-----SGSFVPINGSGTTLSDLSN 456
 DB 424 HKPQGDVLYELGDFNGTILAP--GASTGNIQRLHN 458

RESULT 15
 GUND_CLOCE
 ID GUND_CLOCE STANDARD; PRT; 584 AA.
 AC P25472;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE ENDOGLUCANASE D PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE D)
 DE (CELLULOSE D) (EGGCD).
 GN CELGCD.
 OS Clostridium cellulolyticum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35319;
 RX MEDLINE; 92009193.
 RA Shima S., Igarashi Y., Kodama T.;
 RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
 RT celgcd, of Clostridium cellulolyticum.";
 RL Gene 104:33-38(1991).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODHROLASES THAT CUT THE DISACCHARIDE CELLOBIOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D90341; BAA14354.1; -.
 DR PIR; J01229; J01229.
 DR HSSP; P17901; IEDG.
 DR INTERPRO; IPR001547; -.
 DR INTERPRO; IPR002048; -.
 DR INTERPRO; IPR002105; -.
 DR PRAM; PF00404; Dockerin_1; 2.
 DR PRAM; PF00150; cellulase; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
 DR PROSITE; PS00448; CLOS_CELLULOSEOME_RPT; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 584 ENDOGLUCANASE D.
 FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 329 353 PRO/THR-RICH (LINKER).

FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 530 584 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 530 552 1.
FT REPEAT 562 584 2.
FT ACT_SITE 159 159 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 264 264 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 584 AA: 66061 MW: 0FC41257E81322C3 CRC64;

Query Match 6.5%; Score 170.5; DB 1; Length 584;
Best Local Similarity 20.5%; Pred. No. 0.00017;
Matches 119; Conservative 91; Mismatches 204; Indels 167; Gaps 30;

QY 6 KRIFFSTLLASSILFFSGSTANANGFYVSGTLLYDA-NGNPFVNR---GINHGH- 60
DB 2 KRIALLIISCSIIIMSPFMS-----VYGAINSDQVRRKMGIGMNLGNT 44
QY 61 -----ANYKDAQATAIGIANTGANTVRIYVLDGGQWTKDDIHTVRNLISLAEDNHLV 113
DB 45 FDAPTEGSKAAQAEYFFDFKQAGFKHVRIP-----RW---DQHTLANSPTVDSNFTLN 97
QY 114 AVPEV-----HDATGYDIA-SLNRAVDYWIEMRSALIGKEDTVIINTA 156
DB 98 RIETVIDMELSRGFYVINSHDYMLDMYNSQNGRFEKIMEQIAQRFKRSKNLVFEIL 157
QY 157 NEMFGMEGDADADGKQAIPLRNAGLNHTLAVDAAGWGQFPOSIDYGREYFNADPOR 216
DB 158 NEPHGNTDSQINDMKRLINIRKTNPTRNVIIGAGYNSY-NSLSQI-EIPN-DP-- 211
QY 217 NMFESIHWTE-YAGNAGSOVRTNIDRVNLQDLALVIGFGRHRTNGDVDEATIMSYSEOR 275
DB 212 NLIATPHYIDPISFTHQW-----GTWG---TKNDMD-ALAMAFNHYK 250
QY 276 GVGWLAWSKNGKPEV-----EYIDLSDMA---GNLWAGNTIYNGP 316
DB 251 -----KMSDKNNIPYVLYGEGYGVGHSDDRTSAVWPFVSDQALSHGFCGAMD---NGV 301
QY 317 YGL-----RETSR-----LSTVFTGGSGSGGTSPT-----TLX 344
DB 302 FGSVDNDMAFYNRDTRQDFKEILMALITGTYYDWTPTETNPDPPTPATPAYGEOLIE 361
QY 345 DEGSNQGWTGSSLGCPAVATEWSSKSGSHLKADITQLSSNSOHLVHVIQNTSLOON--- 401
DB 363 DEEGAMQ-W--AAISGVDAATASCISGSKSNNGLETITYAGSSNGTGWVDNEHRNDWEK 418
QY 402 -SRIGATVHANWGSVNGMARTALYKGTGYWYSGSFVPINGSSGTTLSIDLSN--- 456
DB 419 WOKISFDIKSNTNEVRLILAEQSKIEGDEHM--TYIKRSTSMWTIIEIPSSFTKR 475
QY 457 -----VQNLQVREI---GVQFQASDSSGQTSIYIDNV 487
DB 476 MDYQPPADGSETFDLKYKVGSLHFMYSNNSG--TLINDNI 514

Search completed: December 19, 2000, 17:35:51
Job time: 976 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:42:13 ; Search time 248.55 Seconds

(without alignments)
125.108 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607

Sequence: 1 LNNGFKRIFRITSLDASS.....QSASDSSGQTSIYIDNVIYE 490

Scoring table: BLOSUM62

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

P.R.65:*

1: p.r.1:*

2: p.r.2:*

3: p.r.3:*

4: p.r.4:*

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|----------------------|
| 1 | 1614 | 61.9 | 516 | 2 JE0134 | mannan endo-1,4-beta |
| 2 | 761.5 | 29.2 | 384 | 2 T37232 | secreted beta-mann |
| 3 | 731.5 | 28.1 | 363 | 2 S30386 | mannan endo-1,4-beta |
| 4 | 638 | 24.5 | 1331 | 2 A48954 | mannan endo-1,4-beta |
| 5 | 427 | 16.4 | 513 | 2 A37219 | mannan endo-1,4-beta |
| 6 | 239.5 | 9.2 | 557 | 2 JCS487 | cellulase (EC 3.2.2) |
| 7 | 234 | 9.0 | 448 | 2 A27631 | cellulase (EC 3.2.2) |
| 8 | 226.5 | 8.7 | 488 | 2 A25156 | cellulase (EC 3.2.2) |
| 9 | 226 | 8.7 | 409 | 2 B25156 | cellulase (EC 3.2.2) |
| 10 | 220.5 | 8.5 | 505 | 2 S39962 | endo-glucanase |
| 11 | 207 | 7.9 | 504 | 2 S54744 | cellulase (EC 3.2.2) |
| 12 | 203.5 | 7.8 | 570 | 2 S56132 | cellulase (EC 3.2.2) |
| 13 | 193 | 7.4 | 499 | 2 A27198 | cellulase (EC 3.2.2) |
| 14 | 188.5 | 7.2 | 508 | 2 A26874 | cellulase (EC 3.2.2) |
| 15 | 181 | 6.9 | 508 | 2 G69593 | cellulase (EC 3.2.2) |
| 16 | 180 | 6.9 | 486 | 2 T40548 | bi-functional cellu |
| 17 | 179 | 6.9 | 466 | 2 C42360 | cellulase (EC 3.2.2) |
| 18 | 173 | 6.6 | 466 | 2 JCS467 | cellulase (EC 3.2.2) |
| 19 | 171 | 6.6 | 499 | 2 JN0111 | cellulase (EC 3.2.2) |
| 20 | 170.5 | 6.5 | 584 | 2 J01229 | cellulase (EC 3.2.2) |
| 21 | 169 | 6.5 | 822 | 2 J06121 | cellulase (EC 3.2.2) |
| 22 | 168 | 6.4 | 800 | 2 A28003 | cellulase (EC 3.2.2) |
| 23 | 166 | 6.4 | 941 | 2 S28043 | cellulase (EC 3.2.2) |
| 24 | 164.5 | 6.3 | 825 | 2 J50174 | cellulase (EC 3.2.2) |
| 25 | 160.5 | 6.2 | 510 | 2 T37541 | probable glycolipid |
| 26 | 156 | 6.0 | 611 | 2 JCT177 | endo-glucanase V (E |
| 27 | 155 | 5.9 | 814 | 1 C2CLEM | cellulase (EC 3.2.2) |
| 28 | 150 | 5.8 | 357 | 2 PC4404 | cellulase (EC 3.2.2) |
| 29 | 147.5 | 5.7 | 429 | 2 S29044 | endo-glucanase A pr |

| | | | | | |
|----|-------|-----|------|---|----------------------|
| 30 | 146.5 | 5.6 | 747 | 2 | cellulase (EC 3.2.2) |
| 31 | 140.5 | 5.4 | 426 | 2 | cellulase (EC 3.2.2) |
| 32 | 140.5 | 5.4 | 1335 | 2 | glycoprotein VP260 |
| 33 | 135 | 5.2 | 438 | 2 | glucan 1,3-beta-gl |
| 34 | 133 | 5.1 | 1091 | 2 | hypothetical prote |
| 35 | 132.5 | 5.1 | 1608 | 2 | hemolysin A - Ser |
| 36 | 132 | 5.1 | 1300 | 2 | probable serine pr |
| 37 | 131 | 5.0 | 517 | 2 | cellulase (EC 3.2.2) |
| 38 | 130 | 5.0 | 515 | 2 | endo-glucanase - C1 |
| 39 | 129 | 4.9 | 566 | 2 | cellulase (EC 3.2.2) |
| 40 | 129 | 4.9 | 1070 | 2 | cellulase (EC 3.2.2) |
| 41 | 129 | 4.9 | 1536 | 2 | high-molecular-wet |
| 42 | 129 | 4.9 | 2020 | 2 | ABC-type transport |
| 43 | 128.5 | 4.9 | 890 | 2 | ABC-type transport |
| 44 | 127 | 4.9 | 32 | 2 | guar gum-degradin |
| 45 | 126.5 | 4.9 | 428 | 2 | cellulase (EC 3.2.2) |

ALIGNMENTS

RESULT 1
JE0134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N/Alternate names: endo-1,4-beta-mannanase
C/Species: Bacillus circulans
C/Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C/Accession: JE0134
R/Yoshida, S.; Sako, Y.; Uchida, A.
Biosci. Biotechnol. Biochem. 62, 514-520, 1998
A/Title: Cloning, sequence analysis and expression in Escherichia coli of a gene cc
A/Reference number: JE0134; M01D:9623274
A/Molecule type: mRNA
A/Residues: 1-516 <YOS>
A/Cross-references: DDBJ:AB007123
A/Note: the authors translated the codon CAA for residue 259 as Asn. CAA for residue
A/Note: the translation of the nucleotide sequence 1294-1365 is not given in this p
C/Keywords: glycoprotein; glycosidase; hydrolase

| | | | | |
|-----------------------|------------------|-------------------------------------------------------------|------------|-------------|
| Query Match | 61.9% | Score 1614: | DB 2: | Length 516: |
| Best Local Similarity | 59.4%: | Pred. No. 1e-102: | | |
| Matches 291: | Conservative 84: | Mismatches 87: | Indels 28: | Gaps |
| OY | 27 | TSTANANGFYVSGTTLTDANGNPFVARGINGHAMYKDOATTAIEGANTGANTRVYL | 85 | |
| DB | 29 | TNKAAAGFYVSGTKLADATGPPVARGVNHAWYKDOATTAIPALAKGANTRVYL | 88 | |
| OY | 87 | SDGGWTKDIDHTVRNLISLAEDNHLVAVPEVHDATGYDSASLNRAVDYIEMKSLIG | 146 | |
| DB | 89 | ANGHKWTLDPVTVNNITLCEQNKLIVLEVHDATGSDSLDDNNAVYWGISALIG | 148 | |
| OY | 147 | KEDYIINIANEFGSGEDAMADGYKOAIPRLRAGNLHNLVMDAAGMGPDSIHGYC | 206 | |
| DB | 149 | KEDYIINIANEFGTGWGVAMANGYKOAIPRLRAGNLHNLVMDAAGMGPDSVKNYK | 208 | |
| OY | 207 | REVFNAADPQRTMFSIHMYEYAGNASQRTNIDRVLNODIALYIGEGHRTNDVDEA | 266 | |
| DB | 209 | TEVNLADPLKNTVFISIHMYEYAGNASVTKSIDVNLKNLALIIIGEGHRTNDVDEA | 268 | |
| OY | 267 | TIMSYSEGRGVGLAMSKNGKNGPEWEYLDLSNDAGNNLWAGNTIVNGPLRETSRLS | 326 | |
| DB | 269 | TIMSYSEGRGVGLAMSKNGKNGSSDALYLDNTMDAGNSLTSFGNTVNVGSGIKATSYLS | 328 | |
| OY | 327 | TYFTGGSGDGSPTT-----LYDFEGSMGMWTSLSG | 360 | |
| DB | 329 | GIF--GVYTPISSTPTSTPTSTPTSPPTSPGNGNTLLDYEFGTNGMGSGNNISG | 386 | |
| OY | 361 | GPWATYEMSKSHSLKADIOSNSOXYLHVYQTSLOQNSRQATVYKHAMNSVNGM | 420 | |
| DB | 367 | GPWATYEMKATQATLKADVSLQSNSTHSYITTSQNSLGRSSSLKATYKHAMNIGNGI | 446 | |

This Page Blank (uspto)

A: Accession: S30386

A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Keywords: glycosidase; hydrolase; polysaccharide degradation

7

7

7

7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:07:33 ; Search time 362.48 Seconds

(without alignments)
126.223 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNCFKRIFFITSLILASS.....OSASDSSGQSIYIDNYIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 1622 | 62.2 | 516 | 2 | 066185 bacillus ci |
| 2 | 814.5 | 31.2 | 395 | 2 | 069347 vibrio sp. |
| 3 | 761.5 | 29.2 | 384 | 2 | 086599 streptomyces |
| 4 | 634 | 24.3 | 279 | 2 | 092F13 thermomonos |
| 5 | 628 | 24.1 | 930 | 2 | 09RFX5 calidibacill |
| 6 | 541 | 20.8 | 327 | 2 | 09RA15 streptomyces |
| 7 | 253.5 | 9.7 | 1097 | 2 | 09RA17 anaerobic t |
| 8 | 239.5 | 9.2 | 557 | 2 | P94622 clostridium |
| 9 | 228 | 8.7 | 389 | 2 | 059232 bacillus sp |
| 10 | 203.5 | 7.8 | 570 | 2 | 059665 pseudomonas |
| 11 | 193 | 7.4 | 499 | 2 | 045532 bacillus su |
| 12 | 187.5 | 7.2 | 476 | 5 | 018453 heterodera |
| 13 | 187.5 | 7.2 | 481 | 2 | 066064 actinomyces |
| 14 | 187.5 | 7.2 | 501 | 2 | 083012 bacillus sp |
| 15 | 186.5 | 7.2 | 387 | 2 | 031029 erythra car |
| 16 | 185.5 | 7.1 | 930 | 2 | 059290 clostridium |
| 17 | 180 | 6.9 | 486 | 2 | 045430 bacillus sp |
| 18 | 180 | 6.9 | 635 | 2 | 066065 fibrobacter |
| 19 | 179 | 6.9 | 783 | 2 | 045554 bacillus sp |

| | | | | | |
|----|-------|-----|------|----|---------------------|
| 20 | 178 | 6.8 | 749 | 2 | 059154 anaerocellu |
| 21 | 177 | 6.8 | 499 | 2 | 052731 bacillus sp |
| 22 | 173 | 6.6 | 621 | 2 | 007653 cellvibrio |
| 23 | 169 | 6.5 | 821 | 2 | 059241 bacillus sp |
| 24 | 161 | 6.2 | 478 | 5 | 016028 globodera r |
| 25 | 160 | 6.1 | 278 | 2 | 09RKS6 streptomyces |
| 26 | 152 | 5.8 | 506 | 5 | 09UA57 meloidogyne |
| 27 | 148.5 | 5.7 | 470 | 5 | 09UM5 globodera t |
| 28 | 145.5 | 5.6 | 395 | 5 | 09UM4 globodera t |
| 29 | 140.5 | 5.4 | 1335 | 12 | 089353 paramacium |
| 30 | 138.5 | 5.3 | 910 | 3 | P87211 orpomyces |
| 31 | 136.5 | 5.2 | 494 | 2 | 086099 alteromonas |
| 32 | 136 | 5.2 | 438 | 3 | 09UR18 candida alb |
| 33 | 133.5 | 5.1 | 319 | 5 | 018454 heterodera |
| 34 | 132 | 5.1 | 1300 | 2 | 032591 escherichia |
| 35 | 132 | 5.1 | 1300 | 2 | 032555 escherichia |
| 36 | 131.5 | 5.0 | 319 | 5 | 077449 heterodera |
| 37 | 130.5 | 5.0 | 319 | 5 | 061595 heterodera |
| 38 | 130 | 5.0 | 392 | 5 | 044078 globodera r |
| 39 | 129 | 4.9 | 391 | 5 | 077094 globodera r |
| 40 | 129 | 4.9 | 519 | 2 | 047916 fibrobacter |
| 41 | 129 | 4.9 | 1070 | 2 | 055365 synecocyst |
| 42 | 129 | 4.9 | 1107 | 2 | P76088 escherichia |
| 43 | 129 | 4.9 | 1167 | 2 | P76859 escherichia |
| 44 | 129 | 4.9 | 1536 | 2 | 048031 haemophilus |
| 45 | 127.5 | 4.9 | 739 | 2 | 09X687 salmonella |

ALIGNMENTS

| RESULT | ID | PRELIMINARY: | PRT: | 516 AA. |
|--------|----------------------------------------------------------------------|--------------|------|---------|
| 066185 | 066185 | | | |
| AC | 066185 | | | |
| DT | 01-AUG-1998 (TREMBLrel. 07, Created) | | | |
| DT | 01-AUG-1998 (TREMBLrel. 07, Last sequence update) | | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last annotation update) | | | |
| DE | MANNANASE. | | | |
| OS | Bacillus circulans. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| CC | Bacillus/staphylococcus group; Bacillus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE: 96233274. | | | |
| RA | Yoshida S., Sako Y., Uchida A.; | | | |
| RT | "Cloning, sequence analysis, and expression in Escherichia coli of a | | | |
| RT | gene coding for an enzyme from Bacillus circulans K-1 that degrades | | | |
| RT | guar gum." | | | |
| RL | Biosci. Biotechnol. Biochem. 62:514-520(1998). | | | |
| DR | EMBL: AB007123; BAA25878.1; -. | | | |
| DR | INTERPRO: IPR001547; -. | | | |
| DR | PFAM: PF00150; cellulase; 1. | | | |
| SQ | SEQUENCE 516 AA; 55245 MW; DBE5A48842AE8062 CRC64; | | | |

Query Match 62.2%; Score 1622; DB 2; Length 516;
Best Local Similarity 59.6%; Pred. No. 1.2e-107;
Matches 292; Conservative 86; Mismatches 84; Indels 28; Gaps 4;

| | | | |
|----|-----|------------------------------------------------------------|-----|
| QY | 27 | TSTANNSGFYSGITLYANGNPFYMRGINSNGHANYKQATTAIGTANTGYRYLV | 86 |
| DB | 29 | TNNAAHSGFYVSGTKLDATGQPFYMRGVNNAHTYKQDLSTPAIRAKTGNTIRYLV | 88 |
| QY | 87 | SDGQWTKDDITVRNLISLAEDNHLVAPEVDATGYDSIASLRAVDYIEMRSALIG | 146 |
| DB | 89 | ANCHKRTLDVTVNNIILCEONKRIAYLVHDAIGSSLSLDNAVYVIGIKSALIG | 148 |
| QY | 147 | KEDVTIINIANEWGSGWEDANADGYKQAIPLRLNAGLHNTLWDAAGGQFQSIHDYG | 206 |
| DB | 149 | KEDRVIIINIANEWYGTWDAVNAANGYKQAIPLRLNAGLHNTLWDAAGGQFQSIHDYG | 208 |
| QY | 207 | REVFNADPQNRNMFSTHMEFYAGNAGSVRTNIDRYLNDLALVYGFGRHRTNGVDVA | 266 |

| | | | |
|----|-----|---------------------------------------------------------------|-----|
| Db | 209 | TEVNADBLKNTVPSIHXYEYAGNASTVKSINIGVYLNKMLALHIGFEGGQHTNGVDYEA | 268 |
| Qy | 267 | TIMSYSEBRGCGMLAMSKWKGPMWEYIDLSDNPMAGNNLTAMGNTLYNGPYGLRETSRLS | 326 |
| Db | 269 | TIMSTSOQRGCGMLAMSKWKGNSDLAYLDMTNMAGNSLTSPFGTVPVNGSNCGIKATSVLS | 328 |
| Qy | 327 | TYVTFGSGSDGGSPTLT-----LYDFEGSNQMGWMTSG | 360 |
| Db | 329 | GIF--GGVTFPSSPTSTPTSTPTPTPTPTSPGNGTILYDFETGQSGSNNTSG | 366 |
| Qy | 361 | GPMAVTEWSKSGSHSLKADLIOLSSNSOHLVYIONTSLIOONSRIQATVKHANMGVNGM | 420 |
| Db | 387 | GPWVTNEWKATGAQTLKADVSLQSSNTHSLTTSNQMLSGKSLKAYKHAMNGNINCGI | 446 |
| Qy | 421 | TAKLYVKTGHGYIYSGSFVPFINGSGETTSLSDLSNVONLSQVREIGVQFQASDSSGQT | 480 |
| Db | 447 | YAKLYVKVSGMTWYDSGENTLQSDNGTILTLSTSGISNLTSVKEIGVEFPRASSNSGQS | 506 |
| Qy | 481 | SIYIDNVAYE | 490 |
| Db | 507 | AIYDVSYSLQ | 516 |

| | | | | |
|----------------|-------------------------------------------------------------------|-----------------------------|------|---------|
| | RESULT | 2 | | |
| Q69347 | 069347 | | | |
| ID | 069347: | PRELIMINARY: | PRT: | 395 AA. |
| AC | 069347: | | | |
| DT | 01-AUG-1998 (TREMBLE) | 07, Created) | | |
| DT | 01-AUG-1998 (TREMBLE) | 07, Last sequence update) | | |
| DT | 01-MAY-2000 (TREMBLE) | 13, Last annotation update) | | |
| DE | BETA-1,4-MANNANASE. | | | |
| GN | MANA. | | | |
| OS | Vibrio sp. | | | |
| OC | Bacteria; Proteobacteria; | | | |
| RN | [1] gamma subdivision; Vibrionaceae; Vibrio | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-WA-138. | | | |
| RA | Tamatu Y., Arai T., Morishita T., Kimura T., Sakka K., Ohmiya K., | | | |
| RL | J. Ferment. Bioeng. 83:201-205(1997). | | | |
| DR | EMBL; D86329; BAA25188.1; - | | | |
| DR | INTERPRO; IPR001547; - | | | |
| DR | INTERPRO; IPR002883; - | | | |
| PFAM; PR00150; | cellulase; 1. | | | |
| PRAM; PR02013; | CBD_5_2 | | | |
| DR | PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1. | | | |
| SQ | SEQUENCE 395 AA; 43098 MW; A3E30B3A97BCSDAI CRC64; | | | |

| | | | | |
|---------------------------|--------|---------------------|------------|-------------|
| Query Match | 31.28; | Score 814.5; | DB 2; | Length 395; |
| Best Local Similarity | 46.58; | Pred. No. 1.23e-50; | | |
| Matches 168; Conservative | 62; | Mismatches 104; | Indels 27; | Gaps 9 |

[illegible]

```

QY 300 WAGNNLFAAGTIVNGPYGLEETSRLLSTVTTGGG---SPDGTSP---TTLVDEGSMOG 353
      | : : | | : : | : : | : | | | | | | | | | | : : : | : |
Db 291 WDNNYSYTWGAVLLNGQNGIKSTLTATVTCGDCMDCNDBDSSGEPICSSSNAVDPDGDGWC 353
      | : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 353 W 353
      | : : | : : | : : | : : | : : | : : | : : | : : |
Db 351 W 351
      | : : | : : | : : | : : | : : | : : | : : | : : |

```

| RESULT | 3 | | | |
|--------|--------------------------------------------------------------------|-----------------------------------------|------|---------|
| 086599 | | | | |
| ID | 086599 | PRELIMINARY; | PRT; | 384 AA. |
| AC | 086599; | | | |
| DT | 01-NOV-1998 | (TREMBLrel. 08, Created) | | |
| DT | 01-NOV-1998 | (TREMBLrel. 08, Last sequence update) | | |
| DT | 01-MAY-2000 | (TREMBLrel. 13, Last annotation update) | | |
| DE | SECRETED BETA-MANNOSIDASE. | | | |
| GN | MMA2. | | | |
| OS | Streptomyces coelicolor. | | | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinomycetidae; | | | |
| OC | Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-A3(2); | | | |
| RA | Oliver K., Harris D.; | | | |
| RL | Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-A3(2); | | | |
| RA | Parkhill J., Batteil B.G., Rajandream M.A.; | | | |
| RL | Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases. | | | |

RC STRAIN-A5(2);
RX MEDLINE; 97000351.
RA Denenbach M., Kleser H.M., Denapatte D., Eichner A., Cullum J.,
R Kinastil H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb *Streptomyces coelicolor* A5(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031514; CAA20610.1; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002883; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF02013; CBD_5; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 384 AA; 39847 MW; B8AD158E3FD3BD11 CRC64;

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 29.28; | Score 761.5; | DB 2; | Length 384; |
| Best Local Similarity | 44.28; | Pred. NO. 1.3e-46; | | |
| Matches 161; | Conservative 59; | Mismatches 121; | Indels 23; | Gaps 8; |

[illegible]

09RJ15
ID 09RJ15 PRELIMINARY; PRT; 327 AA.
AC 09RJ15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
GN MANA.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser H.M., Denepalte D., Eichner A., Cullum J.,
RA Kinsahl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL121746; CAB57406.1; -.
DR INTERPRO; IPR001547; -.
DR PFAM; PF00150; cellulase; 1.
FT NON_TER 327
SQ SEQUENCE 327 AA; 34772 MW; A21B0068C2C0975D CRC64;

Query Match 20.8%; Score 541; DB 2; Length 327;
Best Local Similarity 41.5%; Pred. No. 5e-31;
Matches 117; Conservative 54; Mismatches 101; Indels 10; Gaps 8;

QY 25 SGTSTANAN-SGFYSGTLLDANGNPFYMRGINHGHAMTKQATTAIGTANTANTYR 83
DB 47 AGPSPGAGAAAGLHGDGLRLENGNDFYMRGVNHAHTYRPE-TQSLADYALGANSVR 105
QY 84 IYLSDGQGTCKDITHTRMILSLAEDNHLVAPEVDATGY--DSTA-SLNRADVWIM 140
DB 106 VVLSGSHRSENGPADVAIVIEGCKANRLICYLEVDTTGYADAAAGTLDAADYIGL 165
QY 141 RSLIGKEDPTVIINANEFWSNEGDAWADGYKQALPRLNAGLNTLWYDAAGWGQFPQ 200
DB 166 KQVLAGEQEDYVIVNIGNEPFGMTDPAGWTEPTVAAVKTLRAAGLQHTINVADAPNMQDQW 225
QY 201 SI-HDYGREVENADPORNMFSIHMEYAGNASOVRTNIDVYNLNDLALVIGEF-GHH 258
DB 226 GYMRANANRKYADPGLNLFISIHMS-VFDTAQETTDVLANFVAELPILIGEGGPAD 284
QY 259 TNGDVEATIMSYSECGYGLWMSKNGGPEMEYDLSDNW 300
DB 285 QYGDPEDEDTMATARGLRGLYGLWMSGNTD--PYDLDALDF 324
RESULT 7
Q9ZAI7 PRELIMINARY; PRT; 1097 AA.
AC Q9ZAI7;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE S-LAYER ASSOCIATED MULTIDOMAIN ENDOGLUCANASE.
GN CELA.
OS anaerobic thermophile KM-THCJ.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
RA [1]

RP SEQUENCE FROM N.A.
RC STRAIN-KM-THCJ;
RX MEDLINE; 99175466.
RA Camm I.K., Kocherginskaya S., King M.R., White B.A., Mackie R.I.;
RT "Molecular cloning, sequencing, and expression of a novel multidomain
RT mannanase gene from the mannanobacterium polysaccharolyticum.";
RL J. Bacteriol. 181:1643-1651(1999).
DR EMBL; U82255; AAD09354.1; -.
DR INTERPRO; IPR001119; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002088; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00395; SH; 3.
DR PFAM; PF02018; CBD_6; 2.
DR PROSITE; PS00904; PPTA; UNKNOWN_1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
SQ SEQUENCE 1097 AA; 119762 MW; C41CEB92C0494B9 CRC64;

Query Match 9.7%; Score 253.5; DB 2; Length 1097;
Best Local Similarity 22.3%; Pred. No. 7.8e-10;
Matches 104; Conservative 76; Mismatches 181; Indels 105; Gaps 20;

QY 8 IFSITSLILASSILFVSGTSTANANSG---FYVSGTLLDANGNPFYMRGIN-HGH-AW 62
DB 10 VFVMTLYALGNGVIGIGKQAKAGTSGDGRFHYGNKIVDPDGNDFYKGVNIQGYRWS 69
QY 63 YKQATTAIEGANTGA-NYVRIYISDG-----GQWTKDDIHTVRNLISLAEDNHLVAV 115
DB 70 EKRSTYADVHLADYWKENTVRLNCFIQNNWMEGTGANNDIDAI--IKAFATKAYVE 126
QY 116 PEVHATGYDSA-----SLNRADVWIMKSLIGMEDYITIANEMGSGW- 164
DB 127 IDLHDTTGPISNPPRPGOSLDOAI-AWFKELAAATYKONPYWENTMNEPSSIAPL 185
QY 165 GDWADGYKQALPRLNAGLNTLWYDAAGWGQFPOSYHD-----YGREVF 210
DB 186 DPQVFAVEEIIKATRSRGADNIIYVD--GWSYANEGIEQNTPTVDEKRSALVYGGDL 243
QY 211 NADPORNMFSIHMEYAGNASOVRTNIDVYNLNDLALVIGEGHRTNGVDATIMS 270
DB 244 NADSAKNTIIFEFHNYN-BGDIQKVEDYIDRANAKGLYFMEVEYKQDS--DAAKEGVKS 300
QY 271 YSE---ORGVGLWMSKNGGPEMEYDLSDNWAGNNTLNGNTIYVPGYLRSTRSLST 327
DB 301 GLQAVMKGAGRIYWNMG---YDLDTLDSG----- 328
QY 328 VFTGGGS-----DQGTPTLLYDFEGSMQGTSSLSGCPWATWESSKSHSLKADIQ 381
DB 329 --TGRSGMEINKTUDGSKPTNL-----SWVGBKIMDNDHGIT--PTFDQPKYDLA 376
QY 382 LSSNQHTLHYQNTSLQNSRIQATYKHAHWGSGVNGMTRKLYK 427
DB 377 LER-----LIANNGFAGDKVQFTFLRNSGDLPIGDKSKVYVK 416
RESULT 8
P94622
AC P94622; PRELIMINARY; PRT; 557 AA.
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE ENDO-1,4-BETA-GLUCANASE ENGf (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN ENGf.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97136706.
RA Shewetta S.A., Ichi-ishi A., Park J.S., Liu C., Malburg L.M.,

Doi R.H.:

RA "Characterization of engf, a gene for a non-cellulosomal Clostridium
RT cellulovorans endoglucanase.";
RL Gene 182:163-167(1996).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
DR EMBL: U37056; AAB40891.1; -.
DR HSSP: O85465; 2A3H.
DR INTERPRO: IPR001547; -.
DR PFM: PF00150; cellulase, 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Hydroxylase; Glycosidase.
SQ SEQUENCE 557 AA; 60131 MW; D186EC88E504EED CRC64;

Query Match 9.2%; Score 239.5; DB 2; Length 557;
Best Local Similarity 23.1%; Pred. No. 3e-09;
Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

QY 2 NNGFKKTSI-----TLSSLLASILEFVSGTS-----TANAN-----SGFTY 38
DB 3 NNVKKKILSIYAAGAMLMALPNNVAAETTSNLTGNANVKKPSVGRKIQLLNNKI-- 60
QY 39 SGTLLDANGNPFVARGIN-HGHAWYKQDATPAIGIANTGA-----NTVRIYL- 86
DB 61 --KILCDKGNFIQLRGKSTHGLQNF-----PGVANNNAFALISNDWNSNVIRLAWY 110
QY 87 -SDGQWMTKDI-HTVRNLSIAEDNHLVAPEVDATGYDSIASLNRAVDYMIERSAL 144
DB 111 VAEGVATNPVSKVQVINGINAIANDMVIYDWMHMGNDNVAISQAQSFNDISLT 170
QY 145 IGKERTYIINIANEWFSGWEG---DA--WADGYKQAIIP---RLRNAGLNLHLMVDAAGW 196
DB 171 YNNKNIITTELCNEPENGNGVTNDATGMAQYKSYATPIVQLLRKNGENLITVGNPFW 230
QY 197 QEPQSHDYGREVFADPQBMTEFISIMYEYAGN-ASQVETNIDR-----VINQDA 248
DB 231 QRPDLAAD-----NPINDSNITSVHF--YSGINPISIVTDNRNANSNVRALHGA 282
QY 249 LVIGEGFHRTNGVDYDEATIMSYSQRGVGLAWSKNGPMEYLDLSND----- 299
DB 283 VFATEMGTSLATG-----TTGPYL-AKADAWLDF-LNGNNISMCNPSISNKEKAALNS 335
QY 300 -----WAGNNLTAMG-----NTIVNGPYGLKRTSLSTLSTVFTGG 332
DB 336 LTSIDPGSKLMADNELTTSQGYRARIKAYATPVPVTQPPAPKDFSGGFDFNDG 395
QY 333 GSDG-----GTSPITLYDFEGSGMQGTSSLSGPPAVTEMSSKSGSHLK-----ADIQLS 383
DB 396 TTQGFGEVNPDSFTITAINVENANNALKISNL-----NSKGSNDLSEGNFMANVRIS 445
QY 384 SNS-OHYLVAVTQNTSLQNSRIQATVKRANMGVNGMTARLYVTGAGY-----TW 434
DB 446 ADIMGOSINITYGDTKLTMDVYAPV-----NVSIAAIPOSSTHGMGPTAIRW 496
QY 435 YSGSFVPINGSSGTTSLDLSNVQNLQVREIGVQFQASDS-----SGQTSYIY 484
DB 497 TNNFVAQTIDGYKATLTISTDPSNFTIA-----TDAASVYVNMILFVSGSDNISTL 550
QY 485 DNV 487
DB 551 DNT 553

RESULT 9
ID 059232 PRELIMINARY; PRT: 389 AA.
AC 059232
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ENDO-BETA-1,4-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULASE)
(ENDOGLUCANASE) (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).

OS *Bacillus* sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC *Bacillus*/Staphylococcus group; *Bacillus*.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-186-1;
RA Sanchez Torres J;
RL Theas (1994), Universidad de Salamanca, Spain.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
DR EMBL: Z33876; CA83942.1; -.
DR HSSP: O85465; 2A3H.
DR INTERPRO: IPR001547; -.
DR PFM: PF00150; cellulase, 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 389 ENDO-1,4-BETA-GLUCANASE.
SQ SEQUENCE 389 AA; 43688 MW; 9111397485609B5F CRC64;

Query Match 8.7%; Score 228; DB 2; Length 389;
Best Local Similarity 23.6%; Pred. No. 1.2e-08;
Matches 100; Conservative 65; Mismatches 154; Indels 104; Gaps 21;

QY 6 KRIEITSLIASILEFVSGTSTANANS-----GFVSGTLLYDANGNPFVNRGI-NH 58
DB 2 KRITTFIVLMTVALFSIGNTTAADNSVVEHQSLSISNGELVNERGEQVQLKGMSSH 61
QY 59 GHAWKQDATPAIGIANTGATVRIYLSDGQWTKD-----IHT----- 99
DB 62 GLQWVG-----QFVNESKMLRDWGITVFPRAMTSSGGYIDPS 103
QY 100 ---YRNLSIAEDNHLVAPEVDATGYDSIASLNRAVDYMIERSALIGKEDYIINI 155
DB 104 VKEKKEAVEALIDDYIYIDMHLSDNDPRITYEKAQFDEH-SELGYGPVNTYI 162
QY 156 ANEWFSGWEGDAMADGYR---QAIPRLRNAGLNLHLMVDAAGWQGFQPSHDYGREVF 211
DB 163 ANEPNPS--DYTWDRIRKPYAEVPIVIRNNDPNIIIV---GTGTSWDVH-HAADNL 216
QY 212 ADPQNTMPSIMYEYAGNQAQVETNIDRVNLQDLAVIGEGFHRHTNGD-----YDEA 266
DB 217 ADP--NVMIAFHF--YAGHGNQLRDQVDYALDQGAALFVSWGISAATGDSGVFLDEAQ 272
QY 267 TMSYSQRGVGLAWS-----WKGNGPMEYLDLSNDWAGNNLTAMGNTIYNG 315
DB 273 EWIDMDERNLSMAMWSLTHKDESSAALMPGANP-----TGTTEALISPSGTFVRE- 324
QY 316 PYGLRETSRLSTVFTGGSDGT-----SPTTLYDFE-----GSM---QGWGSSLSG--- 360
DB 325 --KIRSESISIPSDTPPSDPEYPAAMPQTQYTNELIYVHNGQLQWAKWTONQPGDXY 382
QY 361 GPW 363
DB 383 GPW 385

RESULT 10
ID 059665 PRELIMINARY; PRT: 570 AA.
AC 059665
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGLUCANASE)
(CARBOXYMETHYL CELLULASE).
GN CELL.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
CC Pseudomonas.
RN [1]
SQ SEQUENCE FROM N.A.

RX MEDLINE; 95366948.
 RA Hall J., Black G.W., Ferreira L.M.A., Millward-Sadler S.J.,
 RA All B.R.S., Halsewood G.P., Gilbert H.J.,
 RT "The non-catalytic cellulose-binding domain of a novel cellulase from
 RT Pseudomonas fluorescens subsp. cellulosa is important for the
 RT efficient hydrolysis of Avicel."
 RL Biochem. J. 309:749-756(1995).
 CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL; X86798; CA60493.1; .
 DR HSSP; P07103; 1EG.
 DR INTERPRO; IPR001547; .
 DR INTERPRO; IPR001919; .
 DR INTERPRO; IPR002883; .
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00553; CBD_2; 1.
 DR PFAM; PF02013; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Hydrolase; glycosidase.
 KW SEQUENCE 570 AA; 59925 MW; D00B26F13D57FBE7 CRC64;
 SQ

Query Match 7.8%; Score 203.5; DB 2; Length 570;
 Best Local Similarity 23.3%; Pred. No. 1.1e-06;
 Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

QY 14 SLILASSILFVSGSTANANSGFYVSGTLLDANGNPFVARGIN-----HGHWYKD 65
 DB 22 AIIILGSLGILGVSADADYAPLSTVGAKTIL-RANGOPASFGSLSTWSTWEGEEYNA 80
 QY 66 QATTALEGIANTGANTVRIVL---SDGGQMT---KDIHTVNNILSLAEDNHLVAVPEY 118
 DB 81 QVYSWMLK--SDWAKLVRAANGVEDEGGLTDPANCDRTQYVD-AALINDAVIIDMS 137
 QY 119 HRAIGDSTASLANRANVYIEKRSALIGEDYIINIANEMFG-SWEGADADGKQA-I 176
 DB 138 HNAHQOS-----QATAFQEM-ARKYGANMHIYEIYNEPLQVSVSMT--IKPYAQAVI 189
 QY 177 PRLRNAGLHNTLAVDAAGGQFPQSIHDYGREYFNADP---QRNTFSIHMEYAGNMS 233
 DB 190 AAIKRAIDPNLILGIPPTSSQ-----DVDAANDPITGTONIAYTLHF--YAGTHQO 239
 QY 234 QVRTNIDRYLNDLALVIGFGRHRTNGVDENATIMSYSEORGVGWLAWSKNGPEWEY 293
 DB 240 YLRDAQRLNGLALFVEMGVSANAGD---GAVANSE----- 275
 QY 294 LLSNDMAGNNLTAMGNTIVNGPYGLARETSRLSTVFTGGSGDGSPTLLYDFEGSMQGW 353
 DB 276 ---TNAWSEMT---NHTSNANMALNDKVEGASALVPGASANG-----GW 315
 QY 354 TGSSTL-SGGPMA---VTEW---SSKGSLSLRADIOSNSQHYLHVIONTSIQNSRIQ 405
 DB 316 VNSQLTASALAKSLISGPRSYNTSSSSSAVSSQVSSSQ--APVSSSSSTIASVVS 373
 QY 406 ATV--KHAWSGVNGMTARLYKTGHGYT-----YSGSFVPINGSSTTLLSL 452
 DB 374 SAVSGQCCNM---YGLTYPLCSTTNGWMENNASCIARATCSQGPAPWGIVGSTSS- 428
 QY 453 DLSNVONLSOYVEIGVQFQASDSSGQTSI 482
 DB 429 ---QASSVYSSSSSLVSSSSSSSV 453

RESULT 11
 ID 045532 PRELIMINARY; PRT; 499 AA.
 AC 045532.
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE CELLULASE.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87190397.
 RA Nakamura A., Uozumi T., Teruhiko B.,
 RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis."
 RL Eur. J. Biochem. 164:317-320(1987).
 DR EMBL; M28332; AAA2307.1; .
 DR HSSP; O85465; 2A3H.
 DR INTERPRO; IPR001547; .
 DR INTERPRO; IPR001956; .
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00942; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM; PD001947; .
 SQ SEQUENCE 499 AA; 55075 MW; D424AB9E63B94D23 CRC64;

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 5.2e-06;
 Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

QY 6 KRIFSIITSLILASSILF-----VSGSTANANSG-FYVSGTLLDANGNPFVARG 55
 DB 2 KRISIFITCLILITVLIMGSLQASPASAGKTPRAKKGQSLKGTOLVNDGKAVQLKG 61
 QY 56 I-NCHGANYKQDATTALGIANTGANTVRIYLSDDGGQTKD-----IHT----- 99
 DB 62 ISSHGLQWYGD-----FVNKSLIKWLKRDWIGITVPRAMYTADGGYI 103
 QY 100 -----VRNLISLAEDNHLVAVPEVDATGYSIASLANAVYVIMBSALLGKETV 151
 DB 104 DNPSTKKNVKEAVEAKELGIYIIDWHITLNDGNPNQKKEKAKDFEM--SSLTGNPNV 162
 QY 152 IINIANEMFG--SWEGD--ANADGYKQALPRLRNAGLHNTLAVDAAGGQFPQSIHDYGR 207
 DB 153 IYEINENPNGVNKKRDIKPYAE--EIVSIRKNDPNILIV---GIGTMSQVND--- 213
 QY 208 EYFNADPQ---RNTMFSIHMEYAGNNSQVRTNIDRYLNDLALVIGFGRHRTNG-- 262
 DB 214 ---AADQDKDANWYALHF--YAGTHQOSLRDANALSKGAIFVTEMGTSDASGNG 268
 QY 263 ---VDEATIMSYSEORGVGWLAWSKNGPEW-EYLDLSN--DMAGNNLTAMGNTIVNGPYG 318
 DB 269 VFLQDSR-----EWLNTLDSKNISWVWNL----- 293
 QY 319 LRETSRLSTVFTGGSDGSPPTLLYDFEGSMQGTSSLSGGPWAATWSSKGSLSIKA 378
 DB 294 -----SDKQESSALK-----PGASKTGG-WPLTDLTASGT-FVRE 327
 QY 379 DI--QLSNSQHYLHVIONTSIQNSRIQATVKHANNGSVNGMTARLYKTGHGYT--- 433
 DB 328 NILGNKDKSTKRPETPADNDPAENG--ISVQYKAGDGVNSNQIRPDLHKNNGNATVLD 386
 QY 434 -----WYSGSFVPINGSSTTLLSLDSNY 457
 DB 387 KDVTARYTYNAK-----NKGQNPDCDYAQI 411

RESULT 12
 ID 018453
 AC 018453 PRELIMINARY; PRT; 476 AA.
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE BETA-1,4-ENDOGLUCANASE-1 PRECURSOR (EC 3.2.1.4) (CELLULASE)
 DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).
 GN HG-ENG-1.
 OS Heterodera glycines.
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 98226740.
 RA Smant G., Stockermans J.P.W.G., Yan Y., De Boer J.M., Baum T.J.,
 RA Wang X., Hussey R.S., Gommers F.J., Henriksat B., Davis E.L.,
 RA Helder J., Schots A., Bakker J.;
 RT "Endogenous cellulases in animals: Isolation of beta-1, 4-
 RT endoglucanase genes from two species of plant-parasitic cyst
 RT nematodes";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4906-4911(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-OP50;
 RA Yan Y., Smant G., Stockermans J.W.P.G., Wang X., Hussey R.S.,
 RA Bakker J., Helder J., Schots A., Davis E.L.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: AF006052; AAC48327.1; -
 DR EMBL: AF052733; AAC15707.1; -
 DR HSSP: P07103; 1EG2.
 DR INTERPRO: IPR001347; -
 DR PFM: PF00150; cellulase.1.
 DR SIGNAL: Hydrolase; glycosidase.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 476 BETA-1,4-ENDOLUCANASE-1.
 SQ SEQUENCE 476 AA; 49815 MW; 2056F90CA0783FE7 CRC64;

Query Match 7.2%; Score 187.5; DB 5; Length 476;
 Best Local Similarity 23.7%; Pred. No. 1.2e-05;
 Matches 120; Conservative 61; Mismatches 211; Indels 115; Gaps 23;

QY 13 LSLILASSLITFSG-TSTANANSFGYSGTTLTDANGNPFVWGIN-HGHAWTKD-QAT 68
 DB 13 LFLLLALCALVALYSLTAVALPFGQLSVSGTNLGVANGOPVOLIENLFWHWPQFWMAD 72
 QY 69 TALEGIANTGANTVR--IYLSDGQWTKDDIHTVRL--ISLARDNHLVAPEVHAD 122
 DB 73 TYKALKKNANNAVIRGAMGVDEGGYLS--DANTAYILMAVAIVIAISNGITYLVWHAHN 130
 QY 123 GYDSIASLNRADVWYIEMSSALIGKEDYIINIANEMFG-SWEGDAMADGYKQAIPLRN 181
 DB 131 SHP-----DEAVFEFRIQA-YGSYPHILYEDFNEPLSVSWT-DVLYPHKKVIAAIRA 183
 QY 182 AGLNHTIAMDAGWQ-----FQSIHDYGREFNADPQANTMFSIMHTFYAGNNSQVR 236
 DB 184 IDKKNYILIGTPTWSDVDVAQNPIKDY-----QNLMTYTLHFY-----ASSHP 227
 QY 237 TN-----IDRVLNODLALYIGERG-HRHTNGDVEDATIMSSEQGVGW-LAMSWKNG 288
 DB 228 TNDLGAKLKTAVANNGLPVYETETGCEASGNGNLNDSMSW-----WTLDLSLKIST 280
 QY 289 PEMEYLDLS-----NDMAGNNLTAMGNTIVNGPYGLRSTSLSTVFTGGSD 335
 DB 281 ANMAISDFSEACALSPGTTAAVGVSSRMTSSGNMYAS--YKKKSTGVSCGSSSSGS 338
 QY 336 GGTSPITLIDFESMGWGTGSSLSGGPWATWESKSGSHLAKDIDLSNSOHTLAVION 395
 DB 339 SSSSSSSSSSSSSSSSSSSSSSSG-----SSSSSSSSSSSSSSSSSSSASISVPPN 390
 QY 396 TSLQNSRIQATVKHANMGVYNGMTARLYKTHGYTWSGSFVINGSGSTLSDLS 455
 DB 391 T---WNGGQVNFETKNTISVP-----LCGVVFSVLSPSGTTL----- 425
 QY 456 NVQNTLSQVREIGVQFOSASDSSGQTSI 482
 DB 426 -----GGSNMSESAGSQYSL 441

RESULT 13
 ID 066064 PRELIMINARY; PRT; 481 AA.
 AC 066064;

DT 01-AUG-1998 (TREMREL. 07, Created)
 DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
 DT 01-JUN-2000 (TREMREL. 14, Last annotation update)
 DE ENDOLUCANASE.
 OS Actinomycetes sp. 40.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-40;
 RA Cho K.K.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U94825; AAC06196.1; -
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001347; -
 DR INTERPRO: IPR001919; -
 DR PFM: PF00150; cellulase.1.
 DR PROSITE: PS00659; GLYCOSYL HYDROL. F5; UNKNOWN 1.
 SQ SEQUENCE 481 AA; 50750 MW; 50A6A29BEE44057A CRC64;

Query Match 7.2%; Score 187.5; DB 2; Length 481;
 Best Local Similarity 20.6%; Pred. No. 1.2e-05;
 Matches 100; Conservative 76; Mismatches 212; Indels 97; Gaps 21;

QY 38 VSGTTLTDANGNPFVWGIN-HGHAWTKDQAT-TALEGIANT-GANTVRIVL-----S 87
 DB 45 LSAPTILIDHEGMFQRLGASTHGLQWFGQYVKNKAPQSLDEGLMVRILAVYPRGGYL 104
 QY 88 DGGQWTKDDIHTVRLNISLAEDNHLVAPEVH-----DANGYDSIASLNRADVWYIE 139
 DB 105 QGSQAQMD--KTLGGVQATDGLGMYIIDWHVLYNTPNGDATQAS-----FEK 152
 QY 140 MRSALIGKEDYIINIANEMFSGWEGDAMADGYKQAI-----PRLNAGLNTIAMD 191
 DB 153 SYAAKKSQVNYLFEVCEPTEGT---PYDSGNDIYCYCRMAKAIIDAGSDAILICG 208
 QY 192 AAGWQGFQSIHDYGREFNADPQANTMFSIMHTFYAGNNSQVRTNIDRVLNODLALVY 251
 DB 209 TWTWS---QDIPAVACKPSADGFDNIMVLFH--YAAHCHKDLRAKLTALNAGTPFV 263
 QY 252 GERGF-HRHTNGDVEDATL--MSYSEQGVGLAMSWKNGPEWY-----LDLSDMWG 302
 DB 264 SEFGLDASGNGIDDDSNAMWTLAHNNISTAAALNNKETAAPFPSTVATSKWTG 323
 QY 303 NNTLAMGNTIVNGPYGLRSTSLSTVFTGGSDGTSPTTLXDFESMGWGTG--SSLG 360
 DB 324 DDLTPSAIWLVTNRSKLPADHAAAGTSSGSKASGAGGKKSASASASSPFAAG 383
 QY 361 GPWAV---TEWSSKSGSHLAKDIDLSNSOHTLAVIONTSIQNSRIQATVKHANWGSV 416
 DB 384 GLTATTAIRNNQNSGATYAI--TVSNASGSKH-----EGAMQVTPDLDAEADI-WG-- 432
 QY 417 GNMTRARLYKTHGYTWSGSFVINGSGSTLSDLSNVQNTLSQVREIGVQFOSASDS 476
 DB 433 -----GTIVSHEGTHYVAVAPDMNTALIEGASAEIGFN---ASS 468
 QY 477 SGQTS 481
 DB 469 TGOAS 473

RESULT 14
 ID 083012 PRELIMINARY; PRT; 501 AA.
 AC 083012;
 DT 01-NOV-1998 (TREMREL. 08, Created)
 DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
 DT 01-JUN-2000 (TREMREL. 14, Last annotation update)
 DE CELLULOSE.
 OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-5H;
 RA Khanongnuch C., Ooi T., Kinoshita S.;
 RT "Cloning and nucleotide sequence of beta-mannanase and cellulase gene
 from Bacillus sp. 5H.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016164; BAA31712.1;
 DR HSSP: 085465; 2A3H.
 DR INTERPRO: IPR001547;
 DR INTERPRO: IPR001956;
 DR PFAM: PF00150; cellulase; 1.
 DR PFAM: PF00942; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -; 1.
 SQ SEQUENCE 501 AA; 55422 MW; F76D260A901E2D1D CRC64;

Query Match 7.2%; Score 187.5; DB 2; Length 501;
 Best Local Similarity 21.4%; Pred. No. 1.3e-05;
 Matches 86; Conservative 72; Mismatches 140; Indels 103; Gaps 19;

QY 6 KRTFTSLILASSILFVSGSTANAG-----NSGFVSGTLYDANGNPFVNR 54
 DB 2 KROVSILIVCLLI-SILATSGAGNPVASATGAKTPVALNGQLSIKGTOLINQKAVOLK 60
 QY 55 GI-NHSHAWYKQDQATTAIEGIANTGANTVRIYLSDGQWTKD-----IHT----- 99
 DB 61 GISSHLOWTGD-----FVNRKSLKMLRDMWGLTVPFAAMYTADGGY 102
 QY 100 -----VRNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWIEMSKALIGKEDT 150
 DB 103 IDNPSRKNKKEAVEAKEIGIVIIIDWHILSDNPNYKKEQAKFFAEW-AGLYGNSPTVYETIANEPN 161
 QY 151 VIININAMWFG--SWBGD--AAWDGYKQAIPLRLNAGLNHTLMDVDAAGMGPQSIHDYGR 206
 DB 162 VITEIEEPNGDVNWKRDIPYAE--EVISYIRKNDPNTIIV--GISTWSDQVND-- 213
 QY 207 REVFNADPQ---RNTMFSIMTEYAGGNASQVRIINIDRVLNODLALVIGFGHRTNGD- 262
 DB 214 ----AADDOUKDANVMYALHF--YAGTHGOSLRDKANVALSKGAPFIVETMGSDASNG 267
 QY 263 ----VDEA-TIMSISQDGVGWLAMSKNGPENEYL----DLSDNAGNLTAMGNTIYN 314
 DB 268 GVFLDQSRWLNLTDSKRIKISVMWNLSDKQESSALKPGASKTGWPLTDLTASGIFVRE 327
 QY 315 GPYGLRETSR-----LSTVFTEGGSDGTS 339
 DB 328 NITGNKDSITERPETPAQDNPAQENGISVQYKAG--DGAVN 366

RESULT 15
 ID 031029 PRELIMINARY; PRT; 387 AA.
 AC 031029;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
 GN CELA.
 OS Erwilia carotovora subsp. carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwilia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LY34;
 RA Park Y.W., Lim S.T., Yun H.D.;
 RL Mol. Cells 0:0-0(1997);
 DR EMBL: AF025768; AAC02964.1; -;
 DR HSSP: 085465; 2A3H.
 DR INTERPRO: IPR001547; -;
 DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 KW Signal; Hydrolase.
 FT SIGNAL 1: POTENTIAL.
 FT CHAIN 32 387 BETA(1,4)-GLUCAN GLUCANOHYDROLASE.
 SQ SEQUENCE 387 AA; 42003 MW; 6302B3DF8AC2738B CRC64;

Query Match 7.2%; Score 186.5; DB 2; Length 387;
 Best Local Similarity 21.4%; Pred. No. 1.1e-05;
 Matches 82; Conservative 63; Mismatches 143; Indels 95; Gaps 16;

QY 6 KRTFTSLILASSILFVSGSTANAGS--FYVSGTLYDANGNPFVNRGI-NHSHAWY 63
 DB 10 KRLTGAVTIVGMSLSTALSATPEVETHGQLSINGNLVDEQGRVQLKRVSSHLOWMF 69
 QY 64 KQATTAIEGIANTGANTVRIYLSDGQWTKD-----IHTV 100
 DB 70 GD-----YKDKSMKMLPDDWGINVSVAMYTADGYISKPSLANKV 111
 QY 101 RNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWIEMSKALIGKEDTYINIANEMF 160
 DB 112 KEAVVAAOSIGYIIIDWHILSDNPNYKKEQAKFFAEW-AGLYGNSPTVYETIANEPN 170
 QY 161 G--SWBGD--AAWDGYKQAIPLRLNAGLNHTLMDVDAAGMGPQSIHDYGRVFNADPQR 216
 DB 171 GGVTDGEIRPFA--LEVETIRSKDPDNLIV--GISTWSDQIHD--AADNQLPDP 221
 QY 217 NTFMFSIMTEYAGGNASQVRIINIDRVLNODLALVIGFGHRTNGD-----VDEATISY 271
 DB 222 NTLVYALHF--YAGTHGOFELRIRIDYAOSRGALIFVSEWGTSDASONGGPFLEPESHTEWF 279
 QY 272 SPQDGVGWLAMSKNGPENEYLDLS-----NDWAGNLTAMGNTIYNPGY 318
 DB 280 LNNRG-----GSRVNSILTDKSEASALAPGASKCGWTEDNLSASGKVF----- 324
 QY 319 LRETSRSTVFTEGGSDGTSPT 341
 DB 325 -RAQIRAAATLSGC--DPTTPT 344

Search completed: December 19, 2000, 17:32:55
 Job time: 1522 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:14 ; Search time 226.76 Seconds
(without alignments)
73.889 Million cell updates/sec

Title: US-09-339-159-2
Perfect score: 2607
Sequence: 1 LNNGFKKIRISITSLILASS.....QSAQSDSGGTSTIDNVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.36:*

| | |
|-----|----------------------------------------------|
| 1: | /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:* |
| 2: | /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:* |
| 3: | /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:* |
| 4: | /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:* |
| 5: | /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:* |
| 6: | /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:* |
| 7: | /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:* |
| 8: | /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:* |
| 9: | /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:* |
| 10: | /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:* |
| 11: | /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:* |
| 12: | /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:* |
| 13: | /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:* |
| 14: | /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:* |
| 15: | /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:* |
| 16: | /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:* |
| 17: | /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:* |
| 18: | /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:* |
| 19: | /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:* |
| 20: | /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:* |
| 21: | /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 2607 | 100.0 | 490 | Y54122 | Amino acid sequenc |
| 2 | 1804 | 69.2 | 493 | Y44495 | Bacillus agaradher |
| 3 | 1804 | 69.2 | 493 | Y54124 | Amino acid sequenc |
| 4 | 1711 | 65.6 | 468 | Y44496 | Bacillus agaradher |
| 5 | 1711 | 65.6 | 468 | Y54125 | Amino acid sequenc |
| 6 | 1640 | 62.9 | 476 | Y54123 | Amino acid sequenc |
| 7 | 1152 | 44.2 | 331 | Y54127 | A mannase-linker |
| 8 | 1075 | 41.2 | 369 | Y54129 | Amino acid sequenc |
| 9 | 957.5 | 36.7 | 320 | Y54132 | Amino acid sequenc |
| 10 | 750.5 | 28.8 | 305 | Y54130 | Amino acid sequenc |
| 11 | 546 | 20.9 | 188 | Y54133 | Amino acid sequenc |
| 12 | 419 | 16.1 | 487 | P91007 | Beta-mannase of al |

| | | | | | |
|----|-------|------|------|----|--------|
| 13 | 318 | 12.2 | 132 | 21 | Y54131 |
| 14 | 229.5 | 8.8 | 400 | 18 | W23601 |
| 15 | 229.5 | 8.8 | 400 | 18 | W22521 |
| 16 | 229.5 | 8.8 | 400 | 19 | W57431 |
| 17 | 229.5 | 8.8 | 462 | 19 | W57433 |
| 18 | 226.5 | 8.7 | 410 | 18 | W12378 |
| 19 | 226 | 8.7 | 411 | 18 | W12379 |
| 20 | 224 | 8.6 | 409 | 9 | P81843 |
| 21 | 214 | 8.2 | 1010 | 19 | W34989 |
| 22 | 213 | 8.2 | 467 | 17 | W05731 |
| 23 | 213 | 8.2 | 467 | 17 | W05382 |
| 24 | 211 | 8.1 | 411 | 18 | W12381 |
| 25 | 209.5 | 8.0 | 412 | 18 | W12380 |
| 26 | 193 | 7.4 | 459 | 14 | R42122 |
| 27 | 189 | 7.2 | 551 | 18 | W18790 |
| 28 | 186.5 | 7.2 | 352 | 20 | Y08472 |
| 29 | 179.5 | 6.9 | 476 | 19 | W37241 |
| 30 | 175 | 6.7 | 484 | 19 | W43910 |
| 31 | 169 | 6.5 | 822 | 13 | R26021 |
| 32 | 166 | 6.4 | 941 | 11 | R07478 |
| 33 | 166 | 6.4 | 941 | 16 | R77395 |
| 34 | 164 | 6.3 | 800 | 8 | P70420 |
| 35 | 160 | 6.1 | 531 | 16 | W01503 |
| 36 | 158.5 | 6.1 | 472 | 19 | W37243 |
| 37 | 154.5 | 5.9 | 532 | 12 | R13329 |
| 38 | 151.5 | 5.8 | 562 | 21 | Y69508 |
| 39 | 150 | 5.8 | 357 | 16 | R77394 |
| 40 | 143 | 5.5 | 521 | 17 | R89927 |
| 41 | 142.5 | 5.4 | 358 | 19 | W49874 |
| 42 | 142 | 5.4 | 358 | 19 | W39262 |
| 43 | 129 | 4.9 | 1536 | 14 | R41723 |
| 44 | 129 | 4.9 | 1536 | 14 | R41725 |
| 45 | 129 | 4.9 | 1536 | 15 | R63505 |

ALIGNMENTS

| | |
|--------|--------------------------------------------------------------------------|
| RESULT | 1 |
| ID | Y54122 standard; Protein; 490 AA. |
| XX | Y54122; |
| XX | AC |
| DT | 27-MAR-2000 (first entry) |
| XX | |
| DE | Amino acid sequence of a Bacillus mannanase enzyme. |
| XX | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; |
| KW | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; |
| KW | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; |
| KW | printing paste; plant material degradation; recycled waste paper; |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; |
| KW | mannan-containing food; coffee extract; cleaning composition; |
| KW | machine washing; hard-surface cleaner; dishwashing; oral; dental; |
| KW | contact lens; body-care composition; fabric softener; oil well drilling; |
| KW | subterranean formation fracture. |
| XX | |
| OS | Bacillus sp. |
| XX | |
| EH | Key |
| FT | Peptide |
| FT | 1..30 |
| FT | /note- "signal peptide" |
| FT | 31..330 |
| FT | Domain |
| FT | /note- "catalytic domain" |
| FT | Misc-difference 116 |
| FT | /note- "encoded by CTT" |
| FT | Domain |
| FT | 331..342 |
| FT | /note- "linker" |
| FT | 343..490 |
| FT | Domain |
| FT | /note- "domain of unknown function" |
| XX | |

Amino acid sequenc
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
P300-CelB fusion c
P300-CelB fusion c
Sequence of alkali
Terebintharacter end
Cellulase. Bacill
Bacillus cellulase
P300-CelB fusion c
P300-CelB fusion c
NK-1 cellulase. B
Corrected Bacillus
Actinomycete sp. 3
Heterodera glycine
Heterodera glycine
Alkaline cellulase
Cellulase. Bacill
Full length Bacill
Sequence encoded b
60 kD endoglucanas
Globodera rostochi
Endoglucanase enco
Acidothermus cellu
Fragment of alkali
A. cellulolyticus
Bankia goudi glyc
A. cellulolyticus
High molecular wei
High molecular wei
Haemophilus high m

| | Best Local Similarity | 69.3% | Pred. No. 9.2e-120 | |
|--------|-----------------------|----------------------------------------------------------------------------|--------------------|---------------------------------|
| | Matches, 314 | Conservative | 60 | Mismatches 75; Indels 4; Gaps 3 |
| QY | 11 | IRLSLLSSLSLFVGGSTANANSGFYVSGTLLYANGNPFFMRGINGHAKYKDOATTA | 70 | |
| Db | 12 | llctllslavgmgtc-tpseasctfydgntlllydangqpfwmjgmlnhawyktaasta | 70 | |
| QY | 71 | IEGIANTGANVRIYVLSDDGQWTKDDIHTRNLSIAEDNHLVAVPEVHDATGXDLSASL | 130 | |
| Db | 71 | lpaalaegantlrrlrlvlsdggvkekddltlreivlelaeqnmwvavevhdetgrdsrdl | 130 | |
| QY | 131 | NAAVYVTEMRALLGKRDYIVIANINEMWFGSMEDDANADGKALIPLRNAGLNHTLMV | 190 | |
| Db | 131 | naavylwleemkallgkedvlllnanewysgdswadadyldvlpkrlrdagltlmlv | 190 | |
| QY | 191 | DAAGNGOPFOSLHDYGREVENADPORNTMESIHMYEYAGNNSOVRTNIDRYLNODLALV | 250 | |
| Db | 191 | daagvngqypqslhndyqgdvfnadplkntcmfslmyeyagdnatvtrshldvldgdlaiv | 250 | |
| QY | 251 | IGEFGRHTNGVDDEATIMSYSEORGVGLWMSWKGNGPEWEYLDLSNDWAGNNLTAMGN | 310 | |
| Db | 251 | lgefgrhrtgdgvdtdtllslseeegtylawnswqnsntewdyldlsedwagqhltdcgn | 310 | |
| QY | 311 | TIYNGPFGYLRYSKLSYFT--GGSSDGGTSEPTIITYDEPESNGQGTSSLSGGPAVATEW | 368 | |
| Db | 311 | rlvbgadglgetskspstfytdcddngshpepattllydegstqghvngavtsgpvslew | 370 | |
| QY | 369 | SSKSGHSLKADQLQSSNOHLYHQNTSLQONSLOACTVHANAGSGNGMTALYKTT | 428 | |
| Db | 371 | gssgshslkadnlsnshelysegsnllngysqlnatvhanwgnpnmnarlyvkt | 430 | |
| QY | 429 | GHGTYWTSGSFVPINGS--SGTTLISDLSSNONTL | 460 | |
| Db | 431 | gsdywhagpffrlinsnsqtlstfdlnhienl | 463 | |
| RESULT | 5 | | | |
| ID | Y54125 | standard; Protein; 468 AA. | | |
| XX | XX | Y54125; | | |
| XX | XX | | | |
| DT | 27-MAR-2000 | (first entry) | | |
| DE | | Amino acid sequence of a Bacillus mannanase enzyme. | | |
| XX | XX | | | |
| KW | | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | | |
| KW | | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | | |
| KW | | mannan; galactomannan; glucomannan; galactoglucomannan; cellulostic fibre; | | |
| KW | | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | | |
| KW | | plant material degradation; recycled waste paper; paper making pulp; | | |
| KW | | guar; locust bean gum; mannan-containing food; coffee extract; | | |
| KW | | cleaning composition; machine washing; hard-surface cleaner; | | |
| KW | | dishwashing; oral; dental; contact lens; body-care composition; | | |
| KW | | fabric softener; oil well drilling; subterranean formation fracture;xx | | |
| XX | XX | Synthetic. | | |
| OS | | Bacillus agaradhaerens. | | |
| XX | XX | | | |
| PN | W09964619-A2. | | | |
| XX | XX | | | |
| PD | 16-DEC-1999. | | | |
| XX | XX | | | |
| PF | 10-JUN-1999; | 99WO-DK00314. | | |
| XX | XX | | | |
| PR | 10-JUN-1998; | 98US-0111256. | | |
| PR | 20-OCT-1998; | 98DK-0001340. | | |
| PR | 20-OCT-1998; | 98DK-0001341. | | |
| PR | 28-OCT-1998; | 98US-0105970. | | |
| PR | 28-OCT-1998; | 98US-0106054. | | |
| PR | 23-DEC-1998; | 98DK-0001725. | | |
| PR | 03-MAR-1999; | 99DK-0000306. | | |
| PR | 05-MAR-1999; | 99DK-0000307. | | |

PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Kauppinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
XX WPI: 2000-105891/09.
DR N-PSDB: 245338.
XX
PT New mannanses for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
XX
PS Example 5; Page 215-216; 242pp; English.
XX
CC The present sequence represents a mannanase (also known as mannan
CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The present mannase is a synthetic variant of the mannase of
CC Y54124, in which the C-terminus of the protein was changed due to
CC design of a lower PCR primer used for amplification. The mannase
CC hydrolyses galactomannans. Specifically, mannases hydrolyse
CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulp,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC cellulosics; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.
CC
XX
SQ Sequence 468 AA;
Query Match 65.68; Score 1711; DB 21; Length 468;
Best Local Similarity 69.3%; Pred. No. 9.2e-120;
Matches 314; Conservative 60; Mismatches 75; Indels 4; Gaps 3;
OY 11 ITLSILLASSILFVSGSTANANGFYVSGTIVDANGNPFVMGINGHMYKQATTA 70
DB 12 IICLLISVGMGLT-LSPSAASTGYVDGNTLYDANGPTVMGINDHMYKTAATA 70
OY 71 IEGIANTGANTVRIYLSGSGOMTKDDITVNTLISLADNHLVAPVEYDAGYSISL 130
DB 71 IpaLaegantirivLoeggywekdlditrevlelaeqnmvavevhdatsrsrdl 130
OY 131 NRAVDYIEMKRSALIGKEDYIINANEMFGSGWEGDANADGYKQAIPLRLNAGLHNTLMV 190
DB 131 nravdylmkmkdaligkedvylinaueygsdgsawadgyldvlpkrlrdaegldhltlmv 190
OY 191 DAAGMGEPDSIHGYGRFVNADPQRNMFISIMHYEYAGNAGNSQVTRINDRYLNDLALV 250
DB 191 daagmgpdsihgygrfvnadpqrnmfismhyeyagnagnsqvtrindrylndldlavl 250
OY 251 IGEFGHRTNDVDENATIMSYEQRGVGLAMSMKNGKNGPEYIDLSDMAGNNNTTANGN 310
DB 251 IgefghtndvdenatimsyEQRGVGLAMSMKNGKNGPEYIDLSDMAGNNNTTANGN 310
OY 311 TIYNGPFGRLKRTSLRSLVFT--GGGSDGCTSPPTIYDFEGSMQGTGSLSGGPAVTEW 368
DB 311 tiyngpfgRLKRTSLRSLVFT--GGGSDGCTSPPTIYDFEGSMQGTGSLSGGPAVTEW 368
OY 370 IIVGADGLIGETKSPSTVFTDdngshpepattllyfegstqghgnavvgppsvlew 370
DB 370 IIVGADGLIGETKSPSTVFTDdngshpepattllyfegstqghgnavvgppsvlew 370
OY 428 SSKGSHKADIOJLSSNGHILHYQNTSLQDNRHQATVHANMGSGNGCTALYKXT 428

DB 371 gasngyslkadnvtstnseshlysegerlnhlygqlnatvtrhanwgpngmmarlykt 430
OY 429 GHGTYWISGSFVPINGS-SGTTLSIDLSNQVL 460
DB 431 gsdYtwshgpftrlnssngttlsfdlnnteni 463
RESULT 6
ID Y54123 standard; Protein; 476 AA.
XX Y54123;
AC Y54123;
XX
DT 27-MAR-2000 (first entry)
XX
DE A mannase-linker-cellulose binding domain fusion protein.
XX
KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; Bacillus sp. 1633; galactomannan;
KW 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
KW galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric;
KW printing paste; plant material degradation; recycled waste paper;
KW paper making pulp; guar; locust bean gum; thickener; viscosity;
KW mannan-containing food; coffee extract; cleaning composition;
KW machine washing; hard-surface cleaner; dishwashing; oral; dental;
KW contact lens; body-care composition; fabric softener; oil well drilling;
KW subterranean formation fracture; cellulose binding domain.
OS Synthetic.
OS Bacillus sp.
OS Clostridium thermocellum.
XX
XX WO964619-A2.
XX
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 05-MAR-1999; 99DK-0000309.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Kauppinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX
XX WPI: 2000-105891/09.
DR N-PSDB: 245336.
XX
PT New mannases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
XX
PS Example 4; Page 211-212; 242pp; English.
XX
CC The present sequence represents a mannase-linker-cellulose binding
CC domain fusion protein. Mannase (also known as mannan
CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
CC hydrolyses galactomannans. Specifically, mannases hydrolyse
CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
CC and galactoglucomannans. The mannase protein, or preparations
CC containing it, are used to improve properties of cellulosic or

Db 11 allcsillfagvIntssgaeahbsghvngtlltlydangnptfvmzginhghwfxqet 70

QY 70 AIEGANTGANTVRIVLSDGGQMTKODITHVTNLSLAEONHLVAPPEVHADTGYDSIAS 129

Db 71 smrgtsqgntllrvlndngqrwqkdrnmvasvlslaeqhmlavlevhdatsnfsad 130

QY 130 LNRADVYIEMRSLIGREVTIINIANDEWPGMEGDADQYKQAIPLRARG:GNHTLM 189

Db 131 lgaadvylemdvlygqredvlllnlanewyagwagdgaaqalqlrlnaglshtcm 190

QY 190 VDAAGGQFPOSIDHYGREVENADPQRTMFSIHMYETAGNAGSQRTNIDRVLNQDLAL 249

Db 191 vdaegygyppsvvdygvevlnedpqrntmfsvhmyeyagggancvrrnidsllsqnll 250

QY 250 VIGEGHHTNGDDVDATIMSYSEORGVGWLAMSKNGPEWEYIDLSDMAGNNLTANG 309

Db 251 vlgefghyhdvdedctllsygqgrnvglawshvgnasegydlndfagrnltwq 310

QY 310 NTIYNGPYGLRSTRSLTYF 329

Db 311 dlyvngpqlrtksrsvf 330

RESULT 8

Y54129

ID Y54129 standard; Protein; 369 AA.

XX Y54129;

XX 27-MAR-2000 (first entry)

DE Amino acid sequence of a Bacillus sp. AA349 mannanase enzyme.

XX

KV Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;

KV endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;

KV mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;

KV synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;

KV plant material degradation; recycled waste paper; paper making pulp;

KV guar; locust bean gum; mannan-containing food; coffee extract;

KV cleaning composition; machine washing; hard-surface cleaner;

KV dishwashing; oral; dental; contact lens; body-care composition;

KV fabric softener; oil well drilling; subterranean formation fracture.

XX

OS Bacillus sp.

XX

XX Key

FX Peptide

FT 1..25

FT Location/Qualifiers

FT /note= "signal peptide"

FT 26..67

FT /note= "N-terminal domain of unknown function"

FT 68..369

FT Domain

FT /note= "catalytic domain"

XX

XX W09964619-A2.

XX

PD 16-DEC-1999.

XX

PF 10-JUN-1999; 99WO-DK00314.

XX

PR 10-JUN-1998; 98US-0111256.

PR 20-OCT-1998; 98DK-0001340.

PR 20-OCT-1998; 98DK-0001341.

PR 28-OCT-1998; 98US-0105970.

PR 28-OCT-1998; 98US-0106054.

PR 23-DEC-1998; 98DK-0001725.

PR 05-MAR-1999; 99DK-0000306.

PR 05-MAR-1999; 99DK-0000307.

PR 05-MAR-1999; 99DK-0000308.

PR 05-MAR-1999; 99DK-0000309.

PR 09-MAR-1999; 99US-0123543.

PR 10-MAR-1999; 99US-0123623.

PR 10-MAR-1999; 99US-0123641.

PR 11-MAR-1999; 99US-0123642.

XX (NOVO) NOVO-NORDISK AS.

XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;

PI WPI; 2000-105891/09.

XX N-PSDB; 245342.

DR

XX New mannanases for treatment of textiles, plant material and coffee

PT extract, and in cleaning compositions

PS Claim 35; Page 224-225; 242pp; English.

CC The present sequence represents a mannanase enzyme (also known as

CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).

CC The mannanase hydrolyses galactomannans. Specifically, mannanases

CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,

CC glucomannans, and galactoglucomannans. The mannanase protein, or

CC preparations containing it, are used to improve properties of cellulosic

CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

CC sizes or printing pastes). They are also used to degrade or modify

CC plant materials (particularly recycled waste paper, paper making pulps,

CC or material containing guar or locust bean gums (thickeners), or to

CC reduce viscosity of mannan-containing foods or feeds). The mannanases

CC are also used to process coffee extracts (to inhibit gel formation); in

CC cleaning compositions (for machine washing of fabrics, as hard-surface

CC cleaners, for hand or machine dishwashing, also in oral, dental, contact

CC lens or body-care compositions) where they remove mannan-containing

CC soils and prevent binding of some soils to cellulosics; and in fabric

CC softeners. They can also be used in oil well drilling to fracture

CC subterranean formations.

XX

SO Sequence 369 AA;

Query Match 41.2%; Score 1075; DB 21; Length 369;

Best Local Similarity 66.2%; Pred. No. 1.5e-72;

Matches 196; Conservative 37; Mismatches 63; Indels 0; Gaps 0;

QY 34 SGFYVSGTITLYDANGNPVFMGNGINSHAMWYKQDATTAEIANTGANTVRIYLSGGQMT 93

Db 72 ngfyldgtlltlydangkprtmzginhghwvyrphctamealadtdangsnlrvlsdggqvt 131

QY 94 KODITHVTNLSLAEONHLVAPPEVHADTGYDSIASLNRADVYIEMRSLIGREDTYIT 153

Db 132 kddvdevakllslaekhsivaalevhdaletgtddeplltkdyvleiakdallgkdxkyl 191

QY 134 NIANWPGMEGDADQYKQAIPLRARG:GNHTLMVDAAGGQFPOSIDHYGREVTNAD 213

Db 192 nlsnewfsgvswseggwagdykkaipillreaaglkhtlmvdaagvgqfprshkxglevnsd 251

QY 214 FORNTMFSIHMYETAGNAGSQRTNIDRVLNQDLALVIGEGHHTNGDDVDATIMSYSE 273

Db 252 plktmfmfshyewaaagppqvkndldgvlknlavvlgfghhgygdvavdltlshse 311

QY 274 ORGVGWLAMSKNGPEWEYIDLSDMAGNNLTANGNTIVNGPYGLRSTRSLTYF 329

Db 312 kydvglawshvgnasegydladtdfsgtqltewgerlvhngpqlkxetseivsvy 367

RESULT 9

Y54132

ID Y54132 standard; Protein; 320 AA.

XX Y54132;

XX 27-MAR-2000 (first entry)

DE Amino acid sequence of a Bacillus sp. mannanase enzyme.

XX

KV Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;

KV endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;

KV mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;

KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KM plant material degradation; recycled waste paper; paper making pulp;
 KM guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KM fabric softener; oil well drilling; subterranean formation fracture.
 XX Bacillus sp.
 FT Key
 FT Peptide
 FT Domain
 FT Location/Qualifiers
 FT 1..28
 FT /note= "signal peptide"
 FT 29..320
 FT /note= "catalytic domain"
 PN Mo9964619-A2.
 XX 16-DEC-1999.
 PD 10-JUN-1999;
 PF 99WO-DK00314.
 XX 10-JUN-1998;
 PR 98US-0111256.
 PR 98DK-0001340.
 PR 20-OCT-1998;
 PR 98DK-0001341.
 PR 28-OCT-1998;
 PR 98US-0105970.
 PR 28-OCT-1998;
 PR 98US-0105970.
 PR 23-DEC-1998;
 PR 98DK-0001725.
 PR 05-MAR-1999;
 PR 99DK-0000306.
 PR 05-MAR-1999;
 PR 99DK-0000307.
 PR 05-MAR-1999;
 PR 99DK-0000308.
 PR 05-MAR-1999;
 PR 99DK-0000309.
 PR 09-MAR-1999;
 PR 99US-0123543.
 PR 10-MAR-1999;
 PR 99US-0123623.
 PR 11-MAR-1999;
 PR 99US-0123641.
 PR 99US-0123642.
 XX (NOVO) NOVO-NORDISK AS.
 PA Kaupplinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 PI WPI; 2000-105891/09.
 DR N-PSDB; 245345.
 XX New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PS Claim 35; Page 228-229; 242pp; English.
 XX The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to celluloses; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 SO Sequence 320 AA:

Query Match 36.7%; Score 957.5; DB 21; Length 320;
 Best Local Similarity 58.4%; Pred. No. 7.1e-64;
 Matches 181; Conservative 42; Mismatches 86; Indels 1; Gaps 1;

QY 16 LLAASILFVSGTSTANANGSFYVSGTTLVYDANGNPYMGKINHGAMVYKQDATTAEIGIA 75
 Db 12 LVVVLYIVFVYSSGLAASQSFHVKGTGLIDKNGDPYVMYGVNHGSHVFKGDLSEALPAIA 71
 QY 76 NTGANTRYIYLSDDGOWTKDIDHVRNLISLAEDNHLVAVPEVHDATGYDSIASLNAVD 135
 Db 72 ETGANLTVIVLSNGQWKKDAAELARVIAETGYLTTLVNHDAQSGNPDDLDKAVD 131
 QY 136 YWIMRSALIGKEDTYIINIANEFWGSWEGDAMADGYKQAIPIRLNGLNHTLVADAGV 195
 Db 132 YWIMAEAVLGTEDRVILINANEYGSWRSDEVAEYAGAIPIRLSAGLHTLLVDAAGV 191
 QY 196 GQFQSHIDHREYFNADPQNRMTFSIHMYEXAGNSQVRYNRINDRYLNDLALVIEFG 255
 Db 192 GQYPSLHERGADVFAADPIKNTMFSIMHYEGADRVSENDGYLAENLAVVIEFG 251
 QY 256 HRHTNGVDYEAITMSYSEORGVGLTAMSKGNGEPEWELDISNDWAGNLTAMGNTVING 315
 Db 252 HRHNGVDYEDALITAYTAERGTVGLAWSYGNSSGVGYDLITGPGS-PLTWSGERIVYG 310
 QY 316 PYGLRSTSL 325
 Db 311 EMGLKVIDHL 320
 RESULT 10
 Y54130
 ID Y54130 standard; Protein; 305 AA.
 XX Y54130;
 AC Y54130;
 XX 27-MAR-2000 (first entry)
 DE Amino acid sequence of a partial Bacillus sp. mannanase enzyme.
 XX Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX Bacillus sp.
 OS Mo9964619-A2.
 PN 16-DEC-1999.
 XX 10-JUN-1999;
 PF 99WO-DK00314.
 XX 10-JUN-1998;
 PR 98US-0111256.
 PR 20-OCT-1998;
 PR 98DK-0001340.
 PR 28-OCT-1998;
 PR 98US-0105970.
 PR 28-OCT-1998;
 PR 98US-0105970.
 PR 23-DEC-1998;
 PR 98DK-0001725.
 PR 05-MAR-1999;
 PR 99DK-0000306.
 PR 05-MAR-1999;
 PR 99DK-0000307.
 PR 05-MAR-1999;
 PR 99DK-0000308.
 PR 05-MAR-1999;
 PR 99DK-0000309.
 PR 09-MAR-1999;
 PR 99US-0123543.
 PR 10-MAR-1999;
 PR 99US-0123623.
 PR 11-MAR-1999;
 PR 99US-0123641.
 PR 99US-0123642.
 XX (NOVO) NOVO-NORDISK AS.
 PA Kaupplinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 PI WPI; 2000-105891/09.
 DR N-PSDB; 245345.
 XX New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PS Claim 35; Page 228-229; 242pp; English.
 XX The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to celluloses; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 SO Sequence 320 AA:

CC This protein comprises the alkaline cellulase (endoglucanase) of
CC *Bacillus agaradhaerens* NCIMB 40482. It was expressed in *Bacillus*
CC *subtilis* PL306 transformants following PCR amplification (see
CC T74286-89) of *B. agaradhaerens* genomic DNA and ligation of the PCR
CC product into vector pN1581. DNA encoding the cellulose binding

PI Bjornvad ME, Rasmussen MD, Vind J, Von der Osten C;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:18 ; Search time 154.68 Seconds
(Without alignments)
53.097 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607

Sequence: 1 LNNGRKRIFFSTLSILASS.....QSASDSCGTSITIDNIVE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/PCITUS_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCITUS_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 229.5 | 8.8 | 400 | 2 | US-08-713-298B-2 |
| 2 | 229.5 | 8.8 | 400 | 2 | US-08-870-180B-2 |
| 3 | 229.5 | 8.8 | 400 | 3 | US-08-814-052-4 |
| 4 | 229.5 | 8.8 | 400 | 3 | US-08-812-829-4 |
| 5 | 229.5 | 8.8 | 462 | 2 | US-08-870-180B-13 |
| 6 | 213 | 8.2 | 467 | 2 | US-08-727-548-2 |
| 7 | 189 | 7.2 | 551 | 2 | US-09-033-537A-1 |
| 8 | 174 | 6.7 | 430 | 2 | US-08-924-440-2 |
| 9 | 160 | 6.1 | 531 | 2 | US-07-862-588B-7 |
| 10 | 143 | 5.5 | 521 | 1 | US-08-276-213-3 |
| 11 | 142 | 5.4 | 358 | 1 | US-08-604-913B-11 |
| 12 | 129 | 4.9 | 1536 | 1 | US-08-038-682-2 |
| 13 | 129 | 4.9 | 1536 | 1 | US-08-302-832-2 |
| 14 | 129 | 4.9 | 1536 | 2 | US-08-530-198-2 |
| 15 | 129 | 4.9 | 1536 | 2 | US-08-469-880-2 |
| 16 | 129 | 4.9 | 1536 | 2 | US-08-728-470-2 |
| 17 | 129 | 4.9 | 1536 | 2 | US-08-617-697-2 |
| 18 | 123.5 | 4.7 | 680 | 3 | US-08-947-965-77 |
| 19 | 116.5 | 4.5 | 983 | 3 | US-08-164-292B-26 |
| 20 | 116.5 | 4.5 | 983 | 3 | US-08-845-623-26 |
| 21 | 116.5 | 4.5 | 983 | 3 | US-08-815-927-26 |
| 22 | 112.5 | 4.3 | 480 | 1 | US-08-672-571A-1 |
| 23 | 111 | 4.3 | 574 | 3 | US-08-732-433-1 |
| 24 | 109 | 4.2 | 735 | 2 | US-08-313-185-48 |
| 25 | 109 | 4.2 | 735 | 2 | US-08-459-499-9 |
| 26 | 108 | 4.1 | 1545 | 4 | PCT-US95-10661A-4 |
| 27 | 107.5 | 4.1 | 666 | 2 | US-08-737-716-14 |
| 28 | 107.5 | 4.1 | 1222 | 2 | US-08-682-517-15 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 29 | 107.5 | 4.1 | 1252 | 2 | US-08-682-517-9 | Sequence 9, App11 |
| 30 | 106.5 | 4.1 | 750 | 5 | 5457037-3 | Patent No. 5457037 |
| 31 | 106 | 4.1 | 1338 | 2 | US-08-728-470-9 | Sequence 9, App11 |
| 32 | 106 | 4.1 | 1599 | 2 | US-08-617-697-9 | Sequence 9, App11 |
| 33 | 105.5 | 4.0 | 751 | 5 | 5457037-5 | Patent No. 5457037 |
| 34 | 105 | 4.0 | 385 | 2 | US-08-387-942C-26 | Sequence 26, App1 |
| 35 | 105 | 4.0 | 654 | 1 | US-08-392-828C-2 | Sequence 2, App11 |
| 36 | 105 | 4.0 | 654 | 3 | US-09-330-845-2 | Sequence 2, App11 |
| 37 | 105 | 4.0 | 997 | 2 | US-08-387-942C-4 | Sequence 4, App11 |
| 38 | 104 | 4.0 | 358 | 2 | US-08-853-659A-45 | Sequence 12, App1 |
| 39 | 104 | 4.0 | 735 | 2 | US-08-459-499-12 | Sequence 16, App1 |
| 40 | 103.5 | 4.0 | 553 | 2 | US-08-661-052-16 | Sequence 1, App11 |
| 41 | 103 | 4.0 | 434 | 2 | US-08-989-925-1 | Sequence 5, App11 |
| 42 | 103 | 4.0 | 1702 | 4 | PCT-US95-10661A-5 | Sequence 6, App11 |
| 43 | 102.5 | 3.9 | 623 | 3 | US-09-041-991A-6 | Sequence 4, App11 |
| 44 | 101.5 | 3.9 | 566 | 2 | US-07-862-588B-4 | Sequence 11, App1 |
| 45 | 101.5 | 3.9 | 928 | 1 | US-08-474-140-11 | |

ALIGNMENTS

RESULT 1
US-08-713-298B-2
; Sequence 2, Application US/08713298B
; Patent No. 5922586
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dammann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjernvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5922586 No. 5922586disk of No. 5922586th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713, 298B
; FILING DATE: 13-SEPT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 3794, 424-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 867-0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-713-298B-2

Query Match 8.8%; Score 229.5; DB 2; Length 400;
Best local Similarity 23.4%; Pred. No. 5.8e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

[illegible]

RESULT 2
 US-08-870-180B-2
 Sequence 2, Application US/08870180B
 Patent No. 5945327
 GENERAL INFORMATION:
 APPLICANT: Outtrup, Helle
 APPLICANT: Dammann, Claus
 APPLICANT: Olsen, Arne
 APPLICANT: Bisgaard-Frantzen, Henrik
 APPLICANT: Schjeld, Martin
 APPLICANT: Jrgensen, Per
 APPLICANT: Bjornvad, Mads
 TITLE OF INVENTION: DNA Constructs and Methods of Producing
 TITLE OF INVENTION: Cellulolytic Enzymes
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5945327to No. 5945327disk of No. 5945327th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/870.180B
 FILING DATE: 6-June-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rorek, Carol E.
 REGISTRATION NUMBER: 36,993
 REFERENCE/DOCKET NUMBER: 3794.434-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 400 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-870-180B-2

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 8.8%; | Score 229.5; | DB 2; | Length 400; |
| Best Local Similarity | 23.4%; | Pred. NO. 5.8e-12; | | |
| Matches 97; | Conservative 64; | Mismatches 172; | Indels 81; | Gaps 19; |

```

OY      6 KIKISITILLASSILFVSGTSPAAKN-----GPFVSGTLLYDANGPFWARGI-NH 58
Db      2 KKITTFIVLMTYVALSISGTTTAAODSDSVVEEHGOLSTINGELVNERGQVOIGKMSH 61
OY      59 GHANY-----KDOATTAIEGIANTGANTVATIVE--SPDGQWTDDI-HTVRNL 103
Db      62 GLQWYGVGPVYESKMWLRD-----DMGINVFPAAMYTSSGGVITDDPSVKENYKEA 111
OY      104 ISLAEDNHLVAVPEVHDATGDTSTASLNRAVDYTIEMRSALLCKEDTVIINANEMGWSW 163
Db      112 VEAADIDDIYIYIIMHLLISNDNPPIYKEEKARDFEDEX-SELYGDPYVIEINENPGS- 169
OY      164 EGDAMADGK--QAIPRLRNGLNTLAVDAAGQGFQSTIHDYGREVFNADPQNTM 219
Db      170 -DVTWGNOIKRYAEVETPIIRKNDPNNIITV--GCTWSQDVA-HAADQQLDP--NVM 222
OY      220 FSIHMYEYAGNNAQVYTRINDRYLNODLALVIGEFGRHNGD---VDEATI-MSYSEQ 274
Db      223 YAFHF--YAGPHGNLNDQDYDALDQGAALFVSBWGTSAATGGCGVFLDLDAQWIDPMDE 280
OY      275 RGVCWTLAMS-----WKNGPFMEYELDLSNDMAGNNLTAMGNTIVNGFYGLRETS 333
Db      281 RNLMSWAMSLTHKDESSAALMPCANP-----TGQWTEALSPSGFTVRE--KIRESA 330
OY      324 RLSTVTFGGGSGDGTSPITLYLDFEGSMQGTGSSL-----SGGWANTWEMS 369
Db      331 STPSPDPTPPSDPGEPTPPSPDGEYEPADPNQOITNELIVYHNGQLQMAKWT 384

```

RESULT 3
US-08-814-052-4
Sequence 4, Application US/08814052
Patent No. 6015783
GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Chertty, Joel R.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Vind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
TITLE OF INVENTION: OR STAINS FROM CELLULOSE FABRIC
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60157830 No. 6015783dsk of No. 6015783th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/814,052
FILING DATE: 06-MAR-1997
CLASSIFICATION: 510
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4684.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 400 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-814-052-4

Query Match 8.8%; Score 229.5; DB 3; Length 400;
 Best Local Similarity 23.4%; Pred. No. 5.8e-12;
 Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

6 KRFISITSLILASSILFVSGSTANANS-----GFVSGTLLVDANGNPVWNGI-NH 58
 2 KKITTFVYLLMTVALFSGNTTADNDVSVEHGQLSISNGELVNERGEQYQLKGMSSH 61
 59 GHAVY-----KQATTAEGIANGTAVRVL--SDGGQWTKDDI-HTVRL 103
 62 GLOWGQFVNESMKMLRD-----DMGINFRAMNTSSGGYIDDPVKKKYEKA 111
 104 ISLADNHLVAVPEVHDATGVDISLNRADVWTEMSALIGKEDTYIINIANEFQSW 163
 112 VEAALDILYIYIIDWHILSDNDPNITKEAKDFDEM--SELGDPNVIYEIANEPNGS- 169
 164 EGDADADGK-----QAPRLRAGLNTLMDVADAGQGPQSIHGYGFEVADQQRNTM 219
 170 -DVTGNQIKPYAEVPIPIRNDPNNTIIV--GTGWSQDVH--HADNQLADP--NVM 222
 220 FSIHMYEYAGNAGVQRNIDVRLNODLALVIGFGRHTNGD---VDEATI-MSYSQ 274
 223 YAFHF--YAGTGGVLRQOVYALDQGAALFYSENGTSAITGDGVFLDEAQWIDFDE 280
 275 RGVGLTANS-----WKNGPEWEYLDLSNDAGNNLTANGNTIVNGPYGLRETS 323
 281 RNLSTWANSILTHKDESSAALMPGAMP-----TCGWTAEALSPSGTIVRE--KIRESA 330
 324 RLSYFTGGSGDGTSPITLYDFESSMOGWGSSL-----SGGPAVATWMS 369
 331 SIIPSDPTPPSDPGEPPDTPPSDPEYPAWDPNQIYINELIYVHNQQLWQAKWT 384

RESULT 4
 US-08-812-829-4
 Sequence 4, Application US/08812829
 Patent No. 6017751
 GENERAL INFORMATION:
 APPLICANT: von Oer Osten, Claus
 APPLICANT: Bjornvad, Mads E.
 APPLICANT: Wind, Jesper
 APPLICANT: Rasmussen, Michael Dolberg
 TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 60177510 No. 6017751disk of No. 6017751th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/812.829
 FILING DATE: 06-MAR-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 4690,204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-878-9655
 TELEX:
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 400 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-812-829-4

Query Match 8.8%; Score 229.5; DB 3; Length 400;
 Best Local Similarity 23.4%; Pred. No. 5.8e-12;
 Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

6 KRFISITSLILASSILFVSGSTANANS-----GFVSGTLLVDANGNPVWNGI-NH 58
 2 KKITTFVYLLMTVALFSGNTTADNDVSVEHGQLSISNGELVNERGEQYQLKGMSSH 61
 59 GHAVY-----KQATTAEGIANGTAVRVL--SDGGQWTKDDI-HTVRL 103
 62 GLOWGQFVNESMKMLRD-----DMGINFRAMNTSSGGYIDDPVKKKYEKA 111
 104 ISLADNHLVAVPEVHDATGVDISLNRADVWTEMSALIGKEDTYIINIANEFQSW 163
 112 VEAALDILYIYIIDWHILSDNDPNITKEAKDFDEM--SELGDPNVIYEIANEPNGS- 169
 164 EGDADADGK-----QAPRLRAGLNTLMDVADAGQGPQSIHGYGFEVADQQRNTM 219
 170 -DVTGNQIKPYAEVPIPIRNDPNNTIIV--GTGWSQDVH--HADNQLADP--NVM 222
 220 FSIHMYEYAGNAGVQRNIDVRLNODLALVIGFGRHTNGD---VDEATI-MSYSQ 274
 223 YAFHF--YAGTGGVLRQOVYALDQGAALFYSENGTSAITGDGVFLDEAQWIDFDE 280
 275 RGVGLTANS-----WKNGPEWEYLDLSNDAGNNLTANGNTIVNGPYGLRETS 323
 281 RNLSTWANSILTHKDESSAALMPGAMP-----TCGWTAEALSPSGTIVRE--KIRESA 330
 324 RLSYFTGGSGDGTSPITLYDFESSMOGWGSSL-----SGGPAVATWMS 369
 331 SIIPSDPTPPSDPGEPPDTPPSDPEYPAWDPNQIYINELIYVHNQQLWQAKWT 384

RESULT 5
 US-08-870-180B-13
 Sequence 13, Application US/08870180B
 Patent No. 5945327
 GENERAL INFORMATION:
 APPLICANT: Outtrup, Helle
 APPLICANT: Dammann, Claus
 APPLICANT: Olsen, Arne
 APPLICANT: Bisg rd-Frantzen, Henrik
 APPLICANT: Sch lehn, Martin
 APPLICANT: J rgensen, Per
 APPLICANT: Bjornvad, Mads
 TITLE OF INVENTION: DNA Constructs and Methods of Producing
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 59453270 No. 5945327disk of No. 5945327th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,180B
FILING DATE: 6-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794,434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-180B-13

```

```

Query Match      8.8%; Score 229.5; DB 2; Length 462;
Best Local Similarity 23.4%; Pred. No. 7.3e-12;
Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19;

```

```

QY 6 KIFSTITLLASSILFVSSTANANS-----GFYVSGTLYDANGNPFYMRGI-NH 58
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KITTITFVVLMTALFSGTNTAADNDVVEHGGSLNSELVNERGEVOLKGMSSH 61
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 GHAWY-----KQATTAIEGANTGANTVRIYL--SDGGQWTKDI-HTVRNL 103
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 GLQWYQGFVYVESMKMLRD-----DMGINVFRAMNTSSSGYIDDPSSKEVYKRA 111
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 ISLAEDNHLVAVPEVDATGYDSIASLNAVDWIMRSALIGKEDTVIINANEMFGSM 163
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 112 VEAALIDIDYIILDMHLLSDNDPNIKYKEAKDFDEM--SELYGDYDNYIYEIANEPNGS- 169
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 EGDAMADGK-----QALPRLNAGLHHTLMDVDAAGMGQFPOSIHDPGREYFNA 219
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 -DVTWGNQKPYAEVPIIRNDPNNTIIV--GTGWSQDVH--HAADNQLADP-- 222
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 FSIHMEYAGNAGSVRTNIDRYLNOIDLALVIGFGRHRTNGD---VDPAT--MSYSRQ 274
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 223 YAFHF--YAGTHGQNLRDQVDYALDQGAIFVSEMGTSATGSGVFLEBAQWIDPFMD 280
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 RGVGWLAMS-----WKGNPEWEYLDLSNDWAGNNILFAMGNTIYNGPYGLRETS 323
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 RNLSTANMWSLTHKDESSAALMPGANP-----TGWTEALISPSGTYFRE--KIRESA 330
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 RUSTVTGGSGDGTPTLLYDFEGSMQWTSGL-----SGGPNAVTEMS 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 331 STPPSDPTPPSDGEPDPTPPSDPGKYPAWDPQITNEIYVHNGQLMQAKWMT 384
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
US-08-727-548-2
; Sequence 2, Application US/08727548
; Patent No. 5856165
; GENERAL INFORMATION
; APPLICANT:
; TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/727,548

```

```

CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-727-548-2

```

```

Query Match      8.2%; Score 213; DB 2; Length 467;
Best Local Similarity 22.8%; Pred. No. 2e-10;
Matches 99; Conservative 62; Mismatches 157; Indels 116; Gaps 22;

```

```

QY 6 KIFSTITLLASSILFVSSTSTA-----NANSGFYVSGTLYDANGNPFVMRGI-N 57
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KITTITFVVLMTALFSGTNTAADNDVVEHGGSLNSELVNERGEVOLKGMSSH 60
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 GHAWYKQATTAIEGANTGANTVRIYLSGGQWTKDI-----IHT----- 99
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 HGLQWY-----QFVNYESMKMLRDWGWGITVFRAMNTSSSGYIDDP 102
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 100 -----VRNLISLAEDNHLVAVPEVDATGYDSIASLNAVDWIMRSALIGKEDTVIIN 154
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 SYKEVYKEVEEAALDGIYIILDMHLLSDNDPNIKYKEAKDFDEM--SELYGDYDNYIYE 161
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 INANEMFGS---WEGD--AMADGKQALPRLNAGLHHTLMDVDAAGMGQFPOSIHDPGREV 209
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 INANEMFGSDVTDNDQIKRYAE---EVIPIYIRDNDPNNTIIV---GTGWSQDVH--HAADN 214
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 FNADQRTNMSIHYEYAGNAGSVRTNIDRYLNOIDLALVIGFGRHRTNGD---VDE 265
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 215 QIADP--NVMYAFHF--YAGTHGQNLRDQVDYALDQGAIFVSEMGTSATGSGVFLE 270
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 ATI--MSYSRQGVGWLAMS-----WKGNPEWEYLDLSNDWAGNNILFAMGNTIY 313
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 271 AQWYIDFMDERNLSTANMWSLTHKDESSAALMPGANP-----TGWTEALISPSGTY- 322
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 NGPYGLRETSRLSTVF-----TGGGSDGTSPPTLLYDFEGSMQWTSGL----- 358
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 323 -----REKIRASASITPPSDPTPPSDPGEPDPTPPSD--PEIYRANMSNQIYTN 375
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 ---SGGPNAVTEMS 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 376 VYHNGQLMQAKWMT 389
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION
; APPLICANT: Onishi, Masahito
; APPLICANT: Fitch, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh. Iein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A

```

```

FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492, 204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1

```

```

Query Match          7.2%; Score 189; DB 2; Length 551;
Best Local Similarity 21.3%; Pred. No. 2.9e-08;
Matches 108; Conservative 72; Mismatches 181; Indels 146; Gaps 24;

```

```

QY 38 VSGTLLDANGNPFYMGRI-NHGHWYDQATTAEGIANFGANTVRIYVLSGGGWRD 96
D 11 VQGNQVVGSGQAVQVVGMSHGLQWYGN-----FYKSSIQMMDN 52
QY 97 -IHTVRLSLAEENHLVAVPEVDATGYSIASLNR-----VDYNT----- 138
D 53 WGINVFRAAMTYAEDGY-ITDPYANKKVEAVQASIDGLYIIDWHILSDGNPTTYRAQ 111
QY 139 -----EMKSLIGKEDTYIINANEMFGSWEGDANAD--GY-KQAIPLRLNAGLNTLAV 190
D 112 SKAFQEQATLYGNTPNYIETLANEPNGN---VSNADKSYAEVYITAIRAIDPGGVIV 168
QY 191 DAAGWQFPOSIDHYGREYFNADPQRTMFSIHMYEYAGNAGSOVRTNIDRYLNODIAL 250
D 169 GSPFWSQ-----DIHLAADPVSHSNVYALHF--YSGTHQGFRLDRITYAMNKGALIF 220
QY 251 IGEGRHRTND-----VDEATIMSISQRGVGLAWS-----WNGGPEWEYL 294
D 221 VTEMGTSDASGNGGPFPOSKEWIDFLNARKISWVNMSLADKVENSAALMFGASP----- 275
QY 295 DLSMDMAGNNLTJANGNTIVNGPYGLRETSRLSTVFTGGSDGCT--SPTLLYDEGSMQ 351
D 276 --TGGWIDAOLESSEGRVY-----RQDIROA--TGGSGGNPTIAPAPNTLSATAGNAQ 323
QY 352 -GWTGSSLSGGPMAVTEMSSKSHSLKADIDLSNSOYHLVHTONTLSQNSRIQATYKH 410
D 324 VSLTMNAVSGAT-----SYTVK-----RATISG 346
QY 411 ANMSYSGMGRARLKYKYG--HGYTWY-----SGSFVINGSSGTTSLDLSNVQNLSCV 463
D 347 GPYNNVATGVTATSTYNGTITNGTITTYVVSANS--AGSSANSQAASATPAGSGAST 402
QY 464 REIGVQFO-----SASDSSGQTSIYIDN 486
D 403 GNLVQYKVGDTSATDNOKRPSFNIRKN 429

```

```

RESULT 8
US-08-924-440-2
Sequence 2, Application US/08924440
Patent No. 5871550
GENERAL INFORMATION:
APPLICANT: Frits et al.
TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
NUMBER OF SEQUENCES: 4

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,440
FILING DATE: August 27, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 846-7555
TELEFAX: (650) 846-6504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-924-440-2

```

```

Query Match          6.7%; Score 174; DB 2; Length 430;
Best Local Similarity 24.7%; Pred. No. 4e-07;
Matches 81; Conservative 49; Mismatches 130; Indels 68; Gaps 19;

```

```

QY 26 GTSFNANSG-FYVSGTLLYDANGNPFYMGRI-NHGHWYDQAT-TAIEGTA-NTGANT 81
D 124 GTGPFVERYGQVQVCGIDCEHGNPVQLRGMSTHGIOWFCHCLTDSILDALAYPMKADI 183
QY 82 VRI--VYSDGGWNT--KDDIHTVRLSLAEENHLVAVPEVDATGYSIASLNR 134
D 184 IRLSMTIODEGTEINPGEFTDRM--QLIDMATRGLYIVDWHILTPGPHVYLDRAK 240
QY 135 DWIEMRSLIGKEDTYIINANEMFGSWEGDANAD--GY-KQAIPLRLNAGLNTLAV 191
D 241 TFFAEIARHASKIN-VIETLANE-----PNGVSASIKSYAEVYIPVIRQDDPSVITVG 295
QY 192 AAGWQFPOSIDHYGREYFNADP--QRTMFSIHMYEYAGNAGSOVRTNIDRYLNODIAL 249
D 296 TRGWSLSGV--EGSGPAEIANPVASINMYAFHY-----AASHRDYINLAREASEL 348
QY 250 ---VYGERGHRITND-----VDEATIMSISQRGVGLAWSK----- 285
D 349 PPFVETFGTETTYGDDGANDFQMDR--YIDLMARKIGMTKWNYSDDFRSGAVFQPGTC 406
QY 286 -GNGPEWEYLDLSNDMAGNNLTJANGNTI 312
D 407 ASGGP-----WSSGSLKASGQWY 424

```

```

RESULT 9
US-07-862-588B-7
Sequence 7, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnae
APPLICANT: Sch lein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10017

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/862,588B

FILING DATE: 19920727

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 164/90

FILING DATE: 19-JAN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00013

FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:

NAME: Nelson, Steve T. / Lambart's, Elias J.

REGISTRATION NUMBER: 30,335 / 33,728

REFERENCE/DOCKET NUMBER: 3425,204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 531 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-862-588B-7

Query Match 6.1%; Score 160; DB 2; Length 531;

Best Local Similarity 20.6%; Pred. No. 8.9e-06; Mismatches 109; Conservative 73; Indels 137; Gaps 24;

11 ITTSLIASILFVSGSTANANGS-----FYVSGTLYDANGNPFVNGI-NHGAWTK 64
12 LVLTWVLMGLLPVGPAPKIPAPVPGQATKQGNQVQSGQAVQVQMSHGLQWVG 71
65 DQATTAIEGIANGTANTVRLVLDGGQWTKD--HTVRNLISLADNHLVAVPEVHDAT 122
72 N-----FVNRKSLQWMDNMGINVFRAAMTSEGGY-ITDPSYKNTV 112
123 GYDSIASLNR-----VDYWT-----ENRSALIGKEDVTYINIANEMTG 161
113 KEAVQASIDIALVYIDWHLSDGNPTTYAQSFAFFQEWATLYKNTPNYIYELATSPE 172
162 SWEGDADADYKQAIPLRNAGLNTLMDVDAAGGQFPOSIDHYGREVENADPQRTMS 221
173 CVLGRQSS--EYITFAIRSIDPDGVIVGSPWMSQ-----DHLADAPVSHSNMYA 224
222 IHMYTAGNAGSOVRINIDRLNODLALVYGEHRHTNGDVDEATIMTSSEORGWMLA 281
225 LHF--YSGTGQFLRDIRITAMNGAIFVTEWGTSDASGN----- 283
282 WSKMGNDPEWEYLDLSNDW--AGNNLTANGNTLVNGPYGLRETSRLSTVFGGSDGGT 338
264 -----GGP-----YLPQSKEMIDFLNARKISWVN-----WSLADVEVESALMGASPTGA 309
339 SPTLLYFEGSGMQGTSSLSGGPWAATEWSSKGSLSL-RADIOLSSNSQHYLAVTQNTS 397
310 GP-----MPNCRKSGSSRSN-----PASNMROGNPTAPAPATLSANGCN-----AQVS 355
398 LQONSRLQIA--TYKHAN-----WGSVNGMTARLYKRTG--HEGYTW-----SGSPVPI 442
356 LTMNAVSGATSYTKRAITSSGPGPTNDRGVTAISYNTGLTNGTIIYVVRASNS----- 411

QY 443 NGSSGTTLSLDSNVONLSQVREICVQFO-----SASDSSGGQTSIYIDN 486
DB 412 AGSSANSQAASATPAGSASTGNLVVQYKVGDTSAIDNMRKSFENIKN 459

RESULT 10

US-08-276-213-3

Sequence 3, Application US/08276213

Patent No. 5536655

GENERAL INFORMATION:

APPLICANT: Thomas, Steven

APPLICANT: Laymon, Robert

APPLICANT: Himmel, Michael

TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard

CITY: Golden

STATE: CO

COUNTRY: USA

ZIP: 80401-3393

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/276,213

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: O'Connor, Edna

REGISTRATION NUMBER: 29,252

REFERENCE/DOCKET NUMBER: NREL IR# 94-08

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303)231-1000

TELEFAX: (303)231-1098

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 521 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

US-08-276-213-3

Query Match 5.5%; Score 143; DB 1; Length 521;

Best Local Similarity 21.0%; Pred. No. 0.00025; Mismatches 124; Conservative 70; Indels 216; Gaps 36;

32 ANSGF-YVSGTLYDANGNPFVNGIN-----HGAWTKDQATTAIEGIANAGA 79
1 AGGGYWTHTSGREILDNANVPVRIAGINWGFETCNVYVHG-LMSRDY-RSMIDQIKSLGY 58
80 NTVRIYLSGGQWTKDDI--HTVRNLISLADNHLVAVPEVHDATGVDIASLRADY 136
59 NTRILPYLS-----DILIKPQTMPSINIFYQMN-----QDLQGLTSLQVMDKIYAY 103
137 -----WTEKRSALT-----GREDVTIINIANEW 159
104 AGGIGRITLIDNRDPCSGQSLMTTSSVSATWTISDLQALAQRKGNPTVVGFDLHNE- 162
160 FGSWEGDANADG-----YKQAIPLRNA--GLNHTLAV-----DAAGW----- 195
163 --PHDPACWCGGPDPSIDWHLAERAGNAVLAVNPMLILFVEGVQSYNGSDYWGGLQGA 220
196 GQPP-----QSIDHYGREVFNADPQRTMPSIHMYEYAGNASAV-RVINIDRVLN 244

Db 221 GQPVYLVNPNRYLSADYATSY---PQ--TWSDPTFP---NNMGINKKNGYLEFN 272
QY 245 QDLALV-IGEGFH--RHTNGVDENTINSY---SEQRGVG---WLAWSKNGKNGPEMETLD 295
Db 273 ONIAVFWLGEFGTTLQSTDTQTLVLYLPTAQYAGDSFQWTFWMSNPDSCGTGCI- 331
QY 296 LSNMAGNNLWAKMTYNGPYGLAETSRSLSYFTGGSGSDGTSF--TLYDFBSMGW 353
Db 332 LKDDW-----QTVTV-----KDYLAIDKSIIDPVA----- 360
QY 354 TGSLSGGPMATWESKSHLKADIQLSNSCH-----Y 389
Db 361 -----SASP-----SSQPSVSFSPSPSASRTPTPTPTASPTPTLTPTPTPT 409
QY 390 LHVYQNTSLQNSRIQATVK--HANNGSVNGMTAFLYVKTGHG---YTW-----YSG-- 437
Db 410 ASPSPPTAASGARTASQVNSDW---GNGFTVTVAV--TNSGSVATKWTVTWTFGGNQ 465
QY 438 -----SFVINGSSGTTSLDLSNQNLSQVREIGVFQGSASDSGOT 480
Db 466 TITSMNAAYVQNGQSVTARNKSYNNVIOPGQNTTFGFQASVYGSNAPT 515

RESULT 11

US-08-604-913B-11

Sequence 11, Application US/08604913B

Patent No. 5712142

GENERAL INFORMATION:

APPLICANT: Adney, William S.

APPLICANT: Thomas, Steven R.

APPLICANT: Himmel, Michael E.

APPLICANT: Baker, John O.

APPLICANT: Chou, Yat-Chen

TITLE OF INVENTION: METHOD FOR INCREASING

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: National Renewable Energy Laboratory

STREET: 1617 Cole Boulevard

CITY: Golden

STATE: CO

COUNTRY: U.S.A.

ZIP: 80401-3393

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASC II (DOS) text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/604,913B

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/276,213

FILING DATE: 15-07-1994

ATTORNEY/AGENT INFORMATION:

NAME: Edna M. O'Connor

REGISTRATION NUMBER: 29,252

REFERENCE/DOCKET NUMBER: 95-56

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/384-7573

TELEFAX: 303/384-7459

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 358 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein

FEATURE:

NAME/KEY: E1-CAT

US-08-604-913B-11

Query Match 5.48; Score 142; DB 1; Length 358;
Best Local Similarity 23.18; Pred. No. 0.00017;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 32 ANSGF-VYSGTTVDGANNPFWMGIN-----HGNAWYKDAQTAIEGANTGA 79
Db 1 AGGGTWTSGKEILDANNVPVRIAGINFGFETCNYVHG--LMSHDY-RSMIDQIKSLGY 58
QY 80 NTVRYVLSDDGGQWTKDY---HTVANYLSAEDNHLVAVPEVHDATGYDSIASLRAVDY 136
Db 59 NTRIRPYS-----DLIRKGTMPNSINFYQMN-----QDLQGLTSLQYMDKVIAY 103
QY 137 -----WTEMSALI-----GKEDYIINIANEW 159
Db 104 AGQIGRLIILRHPRDCSGSALMTSSVSSEATWISDQLALQRYKGNPTVAGDILNE- 162
QY 160 FGSWEGDAMADG-----KKAIPRLRNA--GLNHTLV-----DAAGW----- 195
Db 163 --PHDPACWGGGDSIDWRLAERAGNAVLSVNPULLIEGVQSYNGDSYWGNGNLGA 220
QY 196 GQFP-----QSIDYGREYFNADPQRTMESIMYEYAGGNASQV--RTNIDRYLN 244
Db 221 GQPVYLVNPNRYLSADYATSY---PQ--TWSDPTFP---NNMGINKKNGYLEFN 272
QY 245 QDLALV-IGEGFH--RHTNGVDENTINSY---SEQRGVG---WLAWSKNGKNGPEMETLD 288
Db 273 ONIAVFWLGEFGTTLQSTDTQTLVLYLPTAQYAGDSFQWTFWMSNPDSCGTGCI 332
QY 289 -PEWEYLDLSD 299
Db 333 KDDWQTVDTVRD 344

RESULT 12

US-08-038-682-2

Sequence 2, Application US/08038682

Patent No. 5549897

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J

APPLICANT: ST. GEME III, JOSEPH W

TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/038,682

FILING DATE: 16-MAR-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: BERTSTRESSER, JERRY W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-293

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1536 amino acids

TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match 4.98; Score 129; DB 1; Length 1536;
Best Local Similarity 21.18; Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;

```

QY 1 LNNEFKIFSTLSLLASLIFVSGSTANANG-----FYVSGTLLDANGNPEV 52
DB 881 INSG-----NLTAGNTVNIAGNLTVESNAFRAITFTNVGG--LFDKNGSNI 929
QY 53 MRGINHGAWTKDQATTAIEGANTGANTVRIYVSDGGQTKDDIHTVRNLISIAEDNHL 112
DB 930 --STAKGARRKIDNSKNLSITNSSSTYRTIISGNTNNKNGDL---NITNGSDTEM 983
QY 113 VAPEVDATGYDSIAS--LNRAVDYIEMKRSALIGEDYIINIANEMFGSWEGDAMAD 170
DB 984 QIGDVSOKEGNLTISDKIN-----ITKQITKAGVGE-----NSDSD 1023
QY 171 GYKQAIPLRLAAGNLHTLMDVAAGWGPQSIHIDGREVENADPQRTMFSIHMYEYAGG 230
DB 1024 ATNNANLTIKTELKLTODLNISGFNK-----AETAKDSGLTIGTNSAD--GT 1072
QY 231 NASQVRNIDVNLQDIALVIGFGRHTNGVDVDEATIMSYSEORGWGLAMSKGNGPE 290
DB 1073 NAKKV-----TFNQVKRSKISADGKH-----VTLHSKVERTSG-----SNNMTE 1110
QY 291 WEYLDLSNDMAG-----NMLAMGNTIYNGPYGLBETSRSLSTVFTG-----331
DB 1111 ---DSSDNNAGLTIDAKNVTNNITSHKAVSISATSGEITTKGTTINATGNVEITA 1166
QY 332 -----GSGDGTSPPTLYDFEGSMQGTGSSLSGGPVAVTEMS-----KGSHS 375
DB 1167 QTGSLIGIESSSGVTLTATGAL--AVSNISGNTVTVANGALTLTAGSTIKGES 1223
QY 376 LKADIQLS-----SNSOHLVHVIQNTSLQONSRIQATVKAN-----WGSVNGMTARLY 425
DB 1224 VTTSSOSGDIGTISGVEKATESLTQNSKIKATGGEANTVSANGTIGGTISGNTV 1283
QY 426 VATGCTGYWSSGFPVPISSSG-----TTLSDLSNVONLSQVREIGVQFOSASDSGOTS 481
DB 1284 NVTANAAGDLTVNGAELNATGEGATLTTSKGKLTTEASSHTTSANGQVNLASQDSVAGS 1343
QY 482 IYIDNV 487
DB 1344 INAAVY 1349

```

RESULT 13
US-08-302-832-2
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Berenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkesseer, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 4.98; Score 129; DB 1; Length 1536;
Best Local Similarity 21.18; Pred. No. 0.021;
Matches 115; Conservative 74; Mismatches 221; Indels 136; Gaps 21;

```

QY 1 LNNEFKIFSTLSLLASLIFVSGSTANANG-----FYVSGTLLDANGNPEV 52
DB 881 INSG-----NLTAGNTVNIAGNLTVESNAFRAITFTNVGG--LFDKNGSNI 929
QY 53 MRGINHGAWTKDQATTAIEGANTGANTVRIYVSDGGQTKDDIHTVRNLISIAEDNHL 112
DB 930 --STAKGARRKIDNSKNLSITNSSSTYRTIISGNTNNKNGDL---NITNGSDTEM 983
QY 113 VAPEVDATGYDSIAS--LNRAVDYIEMKRSALIGEDYIINIANEMFGSWEGDAMAD 170
DB 984 QIGDVSOKEGNLTISDKIN-----ITKQITKAGVGE-----NSDSD 1023
QY 171 GYKQAIPLRLAAGNLHTLMDVAAGWGPQSIHIDGREVENADPQRTMFSIHMYEYAGG 230
DB 1024 ATNNANLTIKTELKLTODLNISGFNK-----AETAKDSGLTIGTNSAD--GT 1072
QY 231 NASQVRNIDVNLQDIALVIGFGRHTNGVDVDEATIMSYSEORGWGLAMSKGNGPE 290
DB 1073 NAKKV-----TFNQVKRSKISADGKH-----VTLHSKVERTSG-----SNNMTE 1110
QY 291 WEYLDLSNDMAG-----NMLAMGNTIYNGPYGLBETSRSLSTVFTG-----331
DB 1111 ---DSSDNNAGLTIDAKNVTNNITSHKAVSISATSGEITTKGTTINATGNVEITA 1166
QY 332 -----GSGDGTSPPTLYDFEGSMQGTGSSLSGGPVAVTEMS-----KGSHS 375
DB 1167 QTGSLIGIESSSGVTLTATGAL--AVSNISGNTVTVANGALTLTAGSTIKGES 1223
QY 376 LKADIQLS-----SNSOHLVHVIQNTSLQONSRIQATVKAN-----WGSVNGMTARLY 425
DB 1224 VTTSSOSGDIGTISGVEKATESLTQNSKIKATGGEANTVSANGTIGGTISGNTV 1283
QY 426 VATGCTGYWSSGFPVPISSSG-----TTLSDLSNVONLSQVREIGVQFOSASDSGOTS 481
DB 1284 NVTANAAGDLTVNGAELNATGEGATLTTSKGKLTTEASSHTTSANGQVNLASQDSVAGS 1343
QY 482 IYIDNV 487
DB 1344 INAAVY 1349

```

RESULT 14


```
Db 881 INSG-----NLTAGNIVNIAGNLTVESNANFKAITNFTFNVG--LFDNKGNSNI 929
QY 53 MRGINHGAWTKDQATTAIEGIANGTANTVRIVLSDGQMTKDDIHTVRLNISLAEDNHL 112
Db 930 --STAKGARPKDIDNSKNISITNSSSTYRTISGNITKNKNDL---NITNEGSDTEM 983
QY 113 VAPEVHDATGIDSLAS--LRAVDYNTLEMSALIGKEDVYIINIANEFWSMEGDAMAD 170
Db 984 QIGGVVSQKEGNLTISDKIN-----TITKQITIKAGVDE-----NSDSD 1023
QY 171 GYKQAIPLRLRNAGLNIHTLMDVAGNGQFPQSIHDYGREVNADPQRTMFSIHETETAG 230
Db 1024 ATNNANLTKIKELKLTODLINSFNK-----AETAKDSDLLIGNTNSAD--GT 1072
QY 231 NASQVRIIDRYLNODLALVIGFGRHTNGVDDEATIMSISEQRGVGMWAMSNGKNGPE 290
Db 1073 NAKKY-----TFNQVKDSKISADGK-----VTLHSKVEITSG-----SNNTE 1110
QY 291 WEYLDLNDMAG-----NNLTFAMGNTIVNGPYGLRETSRLSTVFTG----- 331
Db 1111 ----DSSDNAGLITDAKNVTYNNNITSHKANVISATSGEITTKGTINATGVETLA 1166
QY 332 -----GGSIDGTSPTTIYDFEGSMQWTSLSGCPNAVTEWS-----KGSHS 375
Db 1167 QTSILGIGIESSSGVTLTATEGAL---AVSNISGNTVTVTANSALTTLAGSTIKTES 1223
QY 376 LKADIQLS-----SNSQHYLVIONTSLOONSRIQATVKHAN---NGSYNGMTARLY 425
Db 1224 VTTSSQSGDIGTISGTVETKATESLITQNSKIKATIGEANVTSACTIGTISGNTV 1283
QY 426 VRTGHGYTWSGFPVINGSSG---TTLISLDSNVQNLISQVREIGVOPQASDSSGQTS 481
Db 1284 NVTANNGDLTVGNGAEINNTGEMALITLTSKGILTEAASHITSAKGOVWLSAQDGSVAGS 1343
QY 482 IYIDNV 487
Db 1344 INANVY 1349
```

Search completed: December 19, 2000, 16:06:50
Job time: 392 sec

Query Match 24.5%, Score 638; DB 2; Length 1331;
 Best Local Similarity 31.4%, Pred. No. 1.4e-35;
 Matches 166; Conservative 84; Mismatches 208; Indels 70; Gaps 14;

```

OY 6 KRIFITSLILASSILFVS-----GTSTANANSGFYSGTLLYDANGNPFYV-----
DB 9 KKMVSILCTVFLNLTILNLTIANVTIIPRYGAAATSN-----DGVYKIDIS-----TLIGN 57
OY 58 HGHANVYKDOATTAIEGIANTGANTVRIYVSDGQWTKDDIHTVRLNLSIAEDNHLVA--VP 116
DB 58 HAHCHYRDRIDTALRGITSMGNSRVYLSNGYRKTIPASVYANILISRLSGRAITL 117
OY 117 EYHDATGT---DSIASLNRADVWIEKMSALIGKEDYIINIANEWFSGWEGDANADGYK 173
DB 118 EYHDGTGEGEDGACSLAQAAYEYKREIKSVLDGNEDEYIINIGNEPYGNNTYQNWNDRK 177
OY 174 QAIPLRANAGLNHTLMDVAAGWQ--FPOSIDHGREYFNADQRTMPSIHMYETAGNA 232
DB 178 NAIKRLRDAEGFHTLVAPNPMGQDSMTMRDNOASIMEADPLNVLVSIMYIG--VYNTA 236
OY 233 SQVFTNIDRVLNQDLATYIEGFRHRTNGDVDEATIMSYSEQRYGWLAMSKNGPEWE 292
DB 237 SKVEYIISFYDKGLPIYIGERHGHOTDGDPEBAIVRYAKYKIGLPSMSKGSSTYVG 296
OY 293 YLDLSNDMAGNNLTAMGNTYNGPYGLRSTRSLSTFTGGSGDGT--SPTLLYDEGSMQ 351
DB 297 YLDVNNMDPNPTPMGQWYKTNAGTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 356
OY 352 GWTGSSLSG-----GPMWTEWSSKSGSHSLKADQLSSNSQRYLHY 393
DB 357 PVSTPATSGQIKVLTYANETNTNTTIRPWL-----RVVNSGSSSIDLSRTYITPTV 410
OY 394 ONTSLOQSRIOATVYKHAMGSGVNGMTARLYKTKG-----GYTWSGSEFVPI 442
DB 411 -----DEBERQSAI--SDMQIGASNTYFKFKVLSSTVSQADYILEIGFKGAQOLDPG 462
OY 443 NGSGTTLISLDSNVQNTSQVREIG--VOFQASDSSGQTSIYIDNVY 489
DB 463 KDTGEIRFNRFDMSYNTNQGNDMSIQMSYGENEKTAYIDGVLY 510

```

RESULT 5
 A37219
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - *Bacillus* sp. (strain AM-001)
 N:Alternate names: 1,4-beta-d-mannan mannohydrolase; beta-mannanase
 C:Species: *Bacillus* sp.
 C>Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 15-Oct-1999
 C:Accession: A37219; B37219; F0111
 R:AKINO, T.; KATO, C.; HORIKOSHI, K.
 Appl. Environ. Microbiol. 55, 3178-3183, 1989
 A:Title: Two *Bacillus* beta-mannanases having different COOH termini are produced in *Esch*
 A:Reference number: A37219; MUID:90146329
 A:Accession: A37219
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-513 <ARI>
 A:Cross-references: GB:M1797; NID:g143166; PIDN:AAA2586.1; PID:g143167
 A:Accession: B37219
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-365 <AK2>
 A:Cross-references: GB:M1797
 R:AKINO, T.; KATO, C.; HORIKOSHI, K.
 Arch. Microbiol. 152, 10-15, 1989
 A:Title: The cloned beta-mannanase gene from alkalophilic *Bacillus* sp. AM-001 produces t
 A:Reference number: F0111
 A:Accession: F0111
 A:Molecule type: protein
 A:Residues: 30-41 <AK3>
 A:Note: three forms of the enzyme (M-I, M-II, and M-III) with different molecular weight
 A:Note: all three forms of the enzyme have the identical amino-terminal sequence shown }

C:Function:
 A:Description: cleaves the beta-1,4-mannosidic linkages in various beta-mannans etc.
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 16.4%, Score 427; DB 2; Length 513;
 Best Local Similarity 26.9%, Pred. No. 9.6e-22;
 Matches 149; Conservative 76; Mismatches 215; Indels 114; Gaps 23;

```

OY 5 EKIFITSLILASSILFVS-----GTSTANANSGFYSGTLLYDANGNPFYV-----
DB 4 YKFAVFAAFIEFSEVLPILISMSSEAN-----GAALSNNANQTKKNYSWLANLPN 55
OY 54 ---KGINHG--AWYKDOATTAIEGIANTGANTVRIYVSDGQW-----TKDDIHTVKN- 102
DB 56 KSNKRVSGHSGEGYSDSTLWAKQCARLGTGMPGILSCDCKNMQTRLYVADQLSYGCNQ 115
OY 103 -LISLAEDNHLVAV-----PEVHDATGYDSIASLNRADV-----WIEKRSALI 145
DB 116 ELINFWNQGGLTTSVHMPPNPFHSGENYKTLPTSQFQNLNHTTGRKKKMDLDMKA 175
OY 146 GKED-----TVIINIANEWFSGWEGDAW--ADGYKQAIPLRANAGLN-----H 186
DB 176 DGLDELQNNQVTVLRLPHENNGEW---FWMGABEYNGFDQTRANAYIASMRDMYQYFTH 232
OY 187 TLMVDAAGWQFPOSIDHYGREYFNADQRTMPSIHMYETAGNASQVRYNIDRVLNQD 246
DB 233 EKKLNLTIMVYSPDYRDH---VTSYPGANYVDVALDST--HPDPSLTLQYNNMIALD 288
OY 247 LALVIGEGHRT--NGDVDEATIMSYSEO---RGYGLAMS--WK--GNGPEWELDISN 298
DB 289 KPPAFALTEGPPESNAGSFDYSYIAIKOKYPRYTFYFLAMDNKSPHNRGAM--DLFN 345
OY 299 DWAGNNLTAMGNTYNGPYGLRSTRSLSTFTGGSGDGT--SPTLLYDEGSMQWGTGS 356
DB 346 D-----SMVYNRGIDYGGSNPAVVLDFENNTLSMGGC 379
OY 357 SLS--GPMWTEWSSKSGSHSLKADQLSSNSQRYLHYQNTSLOQSRIOATVYKHAMGS 415
DB 380 EFTDGPMTSMEWSANGQSILKADVYLVGNNSYH--LQKTYNNRLSFKMLEIKVSHSSWGN 438
OY 416 VGNMTARLYKTKGTYWYSGSFPIGSSGTTSLDSNVQNTSQVREIGYQVQSAUSD 475
DB 439 VSGGTARVYKTSAMWKNAGEFCQFAGKRTALSDLTFSVNLHDVREIGVEYKADAN 498
OY 476 SSGQTSIYIDNVY 489
DB 499 SNGKRAIYIDNVY 512

```

RESULT 6
 JC5487
 cellulase (EC 3.2.1.4) precursor - *Clostridium cellulovorans*
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Clostridium cellulovorans*
 C>Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 15-Oct-1999
 C:Accession: JC5487; PC4333
 R:Shewfelt, S.A.; Ichi-Ishii, A.; Park, J.S.; Liu, C.; Malburg, L.M.; Dol, R.H.
 Gene 182, 163-167, 1996
 A:Title: Characterization of *engF*, a gene for a non-cellulosomal *Clostridium cellulov*
 A:Reference number: JC5487; MUID:97136706
 A:Accession: JC5487
 A:Molecule type: DNA
 A:Residues: 1-557 <SHE1>
 A:Cross-references: GB:U37056; NID:g1778708; PIDN:AAB40891.1; PID:g1778709
 A:Experimental source: strain ATCC 35296
 A:Accession: PC4333
 A:Molecule type: protein
 A:Residues: 30-37 <SHE2>
 C:Comment: This enzyme plays a secondary role in cellulose degradation.
 C:Genetics:
 A:Gene: *engF*
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-53/Product: cellulase #status predicted <CAT>

Query Match 9.2%; Score 239.5; DB 2; Length 557;
Best Local Similarity 23.1%; Pred. No. 6/76-09;
Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

```

QY 2 NNGFKRISI-----TLSLILASSILFVSGTS-----TANAN-----SGRYV 38
DB 3 NNVKKRILISVAGAMALVAVNINVAETYSNLGKNAVKKPSVGKGLQLNKNGI-- 60
QY 39 SGTLLDANGNPFVANGIN-HGHANYKDOATTAIEGIANATGA-----NTRYVL- 86
DB 61 --KTLCDKNGNFIQLRGKSTHLOMF-----PGYVNNNAFALSDMNSNTRYRLMY 110
QY 87 -SDGGQWTKDDI-HYVRNLIISLAEDNHLVAVPEVDATGDSIASLNRAVDYWIEMRSAL 144
DB 111 VAEQGYATNPVYKQFVINGININAIANDMYIVDMHMMNGDNNAVSYGASQSFNDISTL 170
QY 145 IKREDVILINANEMFSGWEG---DA--WADGVKQATP---RLRNAGLNLHTLMDAAGW 196
DB 171 YPNKKNIITELCEPENGEGVTNDATGWAQKSYATPIYQLLRKGNENLITGNPFMS 230
QY 197 OFPQSIHDYGREVFNADPQRTMFSIHMYETAGN-ASQVRNTIDR-----VLNODLA 248
DB 231 QRDGLAAD-----NFINDSMTKSYHF--YSGTNPISYDINRDNAAGNVEYALNHGA 282
QY 249 LVYGEGRHNTNDVDENATIMSYSEQRGVGWLAMSKGKPEMEYIDLSDN----- 299
DB 283 VFATWEGTSLANG---TTGPTL-AKADAMDF--LNGNNISKCNFSINKBEKAAALNS 335
QY 300 -----WAGNMLTAG-----NTIVNGPYGLRRTSRLSTVFTGG 332
DB 336 LTELSDGSKLMDNELITISGOVYRARIKAYATPYDPTNPTAPKFFSSGFMDPNDG 395
QY 333 GSDG---GTSPTLLIDFEGSGMOGWTGSSLSGCPMAVTESSKGSLSK-----ADIOS 383
DB 396 TTQGFQVNPSPITAINVENANNALIKISNL-----NSKGSNDLEGNFMAVNRIS 445
QY 384 SNS-QHLYHIONTSLOONSRIQATYKHAMGWSVNGMTARLYKTHGCV-----TW 434
DB 446 ADIWGOSINITYGDKLMDYIAPPY-----NYSIAIPQSSHGNPTRLIRW 496
QY 435 YSGSFVPIINGSGTLLSDLSNVQNTLSQVREIGVQFOSASDS-----SGQSIYI 484
DB 497 TNNFVAQTDGYKATLITISNDSPNFTIA-----TDADSVYTNMLFVGSNDNISL 550
QY 485 DNV 487
DB 551 DNI 553

```

RESULT 7

A27631
cellulase (EC 3.2.1.4) precursor - Clostridium acetobutylicum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium acetobutylicum
C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
Appl. Environ. Microbiol. 54, 1289-1292, 1988
A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum p2
A:Reference number: A27631; MUID:88268074
A:Accession: A27631
A:Molecule type: DNA
A:Residues: 1-448 <ZAP>
A:Cross-references: EMBL:M31311; NID:g144789; PIDN:AAA23230.1; PID:g144790
A:Note: The authors translated the codon GAG for residue 116 as Gly, GAA for residue 263
C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 9.0%; Score 234; DB 2; Length 448;
Best Local Similarity 22.6%; Pred. No. 1.2e-08;
Matches 115; Conservative 75; Mismatches 185; Indels 134; Gaps 24;

```

QY 3 NGKRTFSTILSLASSILFVSGTSTANANG-----FVSGTLLDANGNPFVNR 54
DB 9 NFKRTFSTILAVM---MLETYLGINTKAEATTSFGQGLKVVGSQCDNSNGRKQLK 65
QY 55 GI-NHGHANY-----KDOATTAIEGIANATVRIVL--SDGGQWTKDDI 97
DB 66 GMSHGLQWTVNPNVNDMSKFLDK-----WGVVYIRAAVNTNNGYISNPSOK 115
QY 98 HYVRNLIISLAEDNHLVAVPEVDATGDSIASLNRAVDYWIEMRSALIKREDVILINAN 157
DB 116 EKIKKIYQDAIDIMVYIIDMHLISDNPNPTYEOKAKSFEQMAEE-YGRYSVITYEION 174
QY 158 EMGSGWGDAMADGK---QATPRLNAGLNLHTLMDAAGWQFQPSIHGYREVFND 213
DB 175 EPNQ---GTNANDIKRYANIITPAIRADPNIIIVGISTWSDVDIAD-----N 223
QY 214 PQR--NTMFSIHMYETAGNAGSOVRINIDRYLNODLALVYGEGRHNTNDVDENATIMSY 271
DB 224 PLRYSNIMYCHF--YAGTHQSLRDKINYAMSGIAIFTEGTSDSAGN----- 272
QY 272 SEQRGVGWLAMSKGKPEMEYIDLSDN---AGNNLTANGNTIVNGPYGLRRTSRLST 327
DB 273 -----GGP---YLDESQKWDVPMASKNI-SWTN-----WALCDKSEASA 307
QY 328 VFTGGSDGSGTSPYLLIDFEGSGMOGWTGSSLS-----GCPMAVTESSKGSLSK 377
DB 308 ALKSSGSGTSG-----GWTDSQTLTGLFVKKISIG-----SWTTSQTS 346
QY 378 ADIOSNSQHYHLYONTSLOONSRIQATYKHAMGWSVNGMTARLYK---TGHGVTW 434
DB 347 PTFISQSTYDQAQTVLITSDNDVHYHTDGTPTSSPVTYISPTIKTTVKAFTT 406
QY 435 YSGSFVPIINGSGTLLSDLSNVQNTLSQVREIGVQFOSASDS-----SGQSIYI 484
DB 407 KTG-----MFDSTITSVNTTISNTDPYKQV 431

```

RESULT 8

A25156
cellulase (EC 3.2.1.4) 1 - Bacillus sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A25156
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic Bacillus sp. s
A:Reference number: A51825; MUID:87056924
A:Accession: A25156
A:Molecule type: DNA
A:Residues: 1-488 <FUR>
A:Cross-references: GB:M4781; GB:X53449; NID:g142659; PIDN:AAA22301.1; PID:g142660
A:Experimental source: strain N-4, plasmid pK1
A:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 8.7%; Score 226.5; DB 2; Length 488;
Best Local Similarity 23.1%; Pred. No. 4.3e-08;
Matches 98; Conservative 67; Mismatches 157; Indels 103; Gaps 22;

```

QY 6 KRIFSITLILASSILFVSGTSTAN-----ANGFYVSGTLLDANGNPFVNRGI-NH 58
DB 2 KRLTTFIYFTLLA--LLFVGNSTSNANGSVVQNGQLSIONQOLVNEHGDVQLKGMSSH 59

```

```

QY   59 GHAUKDQATTAIEGIANTGANTVRIYLSDDGGOWTKRD-----IHT----- 99
Db   60 GLQMG-----OPIANTDISIKMLDDMGITVFRAMTSGGGIYEDBS 101
QY   100 ----VRNLISLAEDNHLVAVPVEDHATGYDSIASINRAVDYWIEMKSALIGKEDTYINI 155
Db   102 VKEKKEAVEAIDIGIYIIIDMHILSDNDPNIYKEEAKEFFDEM-SALYGYPNTIYEI 160
QY   156 ANEWFG---SWEGDMADGY-KQAIPRLRNAGLNHTLTMYDAAGKGOFPOSIHDTGREYFN 211
Db   161 ANEPNGHNWVR--DSHIRPYAEVPIPIRANPNNVIVIGTATWS---QDVHEADNOLD 215
QY   212 ADPOQNTFESIMTYAYAGNSAQVFRNIDRYLNOIDLALYIGEGRHRTGD----VDENT 267
Db   216 -DP-NVATIAEF--YAGTHGQLRNOVDYALSRCALAIIVSSWGTSAAATGDGGVFLDEAQ 270
QY   268 I-MSYSEORGVOGLAWS-----WKNGPEWEYLDLINDNAGNYLTAMGNTIYVG 315
Db   271 VWIDPMDBERNLSMAWMSLTHNDESSAALLPGANP-----TGWTFAELSTSGAFVRE- 322
QY   316 PYGLRETRSLSTVFTGGSD--GGTSPTTLTYDFGSMOGWTGSSL-----SGGPWA 364
Db   323 --KRRESASIPSPDPTPPSDPDGPEDPTPPSD-PGEYPAAMPNOIYTNEIYHNGQLMQ 379
QY   365 VTEMS 369
Db   380 AKMWT 384

```

RESULT 9
B25156
cellulase (EC 3.2.1.4) 2 - *Bacillus* sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: *Bacillus* sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: B25156
R:Fukumori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. *streus*
A:Reference number: A91825; MIMD:87056924
A:Accession: B25156
A:Molecule type: DNA
A:Residues: 1,409 <FUNK>
A:Cross-references: GB:M14729; NID:g142655; PIDN:AAA2229.1; PID:g142656
A:Experimental source: strain N-4, plasmid pNK2
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
:Keywords: glycosidase; hydrolase; polysaccharide degradation

[illegible]

```

Db 215 QUTDP--VMXAFH--YAGHGNQLRQVDYDALDQSAALFVSWMGSEATGGGVFLDE 270
QY 266 ATI-MSYEEGRCGVGNLAWS-----WKGNGPMEYELDISNDWAGNNLTAMGNTIV 313
Db 271 AQWVIDPFMDERLISANNSTLHKDESSAALMPGASP-----TGCTWEALSPSGTFV- 322
QY 314 NGPYGLRRTSRLSYFTGG-----SDGCT-----SPTLYD 345
Db 323 -----RKIRISATPPSDPTPPSDPDGPEHPDGPEDPTPPSDPDGDPAMDPTIYT 376
QY 346 FE-----GSW---OGWTGSSLSG---GPW 363
Db 377 DEIVYHNGQLMQAKWNTIONQSPGDPYGFW 405

```

```

RESULT 10
S39962
endoglucanase - Erythrina carotovora
C:Species: Erythrina carotovora
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S39962
R:Cooper, V.J.C.; Salmond, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A:Title: Molecular analysis of the major cellulase (CelV) of Erythrina carotovora: evid
A:Reference number: S39962; MUID:94067016
A:Accession: S39962
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1505 <COO>
A:Cross-references: EMBL:X76000; NTD:9434941; PIDD:CAM53592.1; PID:9434942

```

| Query Match | 8.5% | Score 220.5 | DB 2 | Length 505 |
|------------------|--------------|-------------------------------------------------------------|----------------|------------|
| Local Similarity | 21.8% | Pred. 1.12e-07 | | |
| Matches 116 | Conservative | 69 | Mismatches 209 | Indels 137 |
| | | | | Gaps 21 |
| QY | 2 | NNGFKRISITLLSLIASILFVSGSTANANSG-FYVSGTLLDANGNPEYMRGI-NHG | 59 | |
| DB | 6 | NQIVKRLTGVVTVYLGMSLSPSALSAPFVETRHGQLSTENGRVLDEQGRVQLRGISSHG | 65 | |
| QY | 60 | HAWYKDDATTAEGIANGTAVRYVLSDDGGWTDD- | 96 | |
| DB | 66 | LOWFGD-----VYNKDSMKWLDDDMGIVFRVAMYTADGYISNPL | 107 | |
| QY | 97 | IHTVNLISLADNHNVAPEYHATGVDYSIASLRADVYTEMRSALIGKEDTYIINIA | 156 | |
| DB | 108 | AKKVEAANAASQSLGYITIIDWHLISDNDPNITYKAOKATFEEM-AGLGSSPNYIYETA | 166 | |
| QY | 157 | NEMFG--SWEGD--AWADGYKQAIPLRNAIGHTLMDYDAAGWQGFPOSIDHGREVENA | 212 | |
| DB | 167 | NEPNGVMTNGGIRRYA---LEVDTIRSKDPDNLIIY--GTGRWSDIHD--AADNQ | 217 | |
| QY | 213 | DPQRTMSTIHETVAGGNASOVFTNIDRVLNODIALYIGEPGHRHTNGD---- | 267 | |
| DB | 218 | IEDPPTMYALHF--YAGTEGQFLRIRIDYAOGRGAIFVSEWGTSDASGNGGPTLPESQT | 275 | |
| QY | 268 | IMSSEQRGVGLAMSKXGNPMEYL-----DLSMDMGNNTLTCNGNTVYNPGIARETS | 323 | |
| DB | 276 | WIDFLNNRVSVMWNNSLTDKSEASALALPAGAKSGGWTEQNLSTGKFEV----- | 328 | |
| QY | 324 | RLSTFTGGGSGGCTSPFTLLYDFEGSGMGWTSLSGGFWATYEWSSKGSLSIKADIOLS | 383 | |
| DB | 329 | R-----AGNAGGCGDPTTPTEPTPGKNTTC-----DVLV- | 359 | |
| QY | 384 | SNSQHTLAVYQNTSLQOONSRIQATYKHANWGSVG--NGMTARLYKTCGTYWSSGFV | 440 | |
| DB | 360 | -----GYRNDNPNPS--DDAIRMAVINIKNTGSTPIKSLDLYQRYFHD----- | 400 | |
| QY | 441 | PINGSGGTLLSLDLSNVQLTSQYREIGVQFQASD-----SSGQTSI | 482 | |
| DB | 401 | ---DGRPGANLEFDMANVGNPNVITSTGTPPAASTKANRIVYLVTFSSGGSL | 449 | |

RESULT 11

S54744 cellulase (EC 3.2.1.4) CelV precursor - *Erythrina carotovora* (SCC 3193)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Erythrina carotovora*

A:Variety: SCC 3193

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S54744; S44996

R:Maee, A.; Heikkinen, R.; Palva, E.T.

Mol. Gen. Genet. 247, 17-26, 1995

A:Title: Structure and regulation of the *Erythrina carotovora* subspecies *carotovora* SCC3193

A:Reference number: S54744; MUID:95231512

A:Accession: S54744

A:Molecule type: DNA

A:Residues: 1-504 <MAX>

A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g493493

A:Genetics:

A:Gene: celV1

C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-33/Domain: signal sequence #status predicted <SIG>

F:33-504/Product: cellulase #status predicted <MAT>

Query Match

Best Local Similarity 7.9%; Score 207; DB 2; Length 504;

Matches 110; Conservative 63; Mismatches 191; Indels 130; Gaps 22;

QY 6 KRIFSTLSLLASSTLPVSGTSTANANG-FTVSGTLYDANGPFWNRGI-NHGHMY 63
 DB 10 RKLTLLGVYTMGLMSFSLSATPVETHGQSLIENGRLVDQGRKVOGLSSNLOW 69
 QY 64 KQATTAIGTANTGANTVRIYLSGGQWTKDD-THFRNLISLAEDNHLVAPEV 118
 DB 70 GD-----YVKKDSMKLRLDWMGINFRVAMTYAE-NGYLANSLANK 110
 QY 119 -----HDATGYDSIASLNRADVWTEKRSALIGKEDTYIINANEM 159
 DB 111 VKFAVAAAGLGYIILIDHTLSNDPNTYKAQAKIFFAEM-AGLGNPNVITYETANEP 169
 QY 160 FGS--WEGD--AMADGYKALPRLNAGLNHTLVDAAGCGFPOSIDHYGREVFADQ 215
 DB 170 NGSVTMNGQIRPYA--LEVTDTIRSDPNLIIYSGSTVS--QDIHD--AADQOLD 220
 QY 216 RNTMFSIHMEYVAGNAGQVRINIDRYLNQDLALVIGFGRHTND-----VDEATIMS 270
 DB 221 PNTLYALHF--YAGTHGQFLRDRIDYAGSRGALIFVSEMGTSASGNGGFLPEESQTWID 278
 QY 271 YSEORGVGLAWSKNGKGPWEYL-----DISNDMAGNNLTAMGNTYNGYGLRETSL 326
 DB 279 FLNKRGISVNWMSLSDKSETSAALVAGASKSGGTEONLSISGFV-----REQIR 329
 QY 327 TVFTGGSDGSPPTLYLDFEGSMQGTGSLSGFPWAVTEWSKSGSLKADIOSNS 386
 DB 330 ---AGALSGGDTPTMTEPTNPNGTGT-----DIVL----- 359
 QY 387 QHLYLVQNTSLQONSRIQATVKAHNGSVG--NGATPALLYKTGTGTYTSGSEFPIN 443
 DB 360 -QYRNVNPNPS---DDAIRMAFNKIKNGTSPFKLSLDQVRYFHD-----D 401
 QY 444 GSSGTTLSDLSNY 457
 DB 402 GKPGANLEVDWANY 415

RESULT 12

S56132 cellulase (EC 3.2.1.4) precursor - *Pseudomonas fluorescens*

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Pseudomonas fluorescens*

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56132; S56127; S54798

R:Hall, J.; Black, G.W.; Ferreira, L.M.A.; Millward-Sadler, S.J.; All, B.R.S.; Hazlewood

Biochem. J. 309, 749-756, 1995

A:Title: The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas*

A:Reference number: S56127; MUID:95366948

A:Accession: S56132

A:Molecule type: DNA

A:Residues: 1-570 <MAX>

A:Cross-references: EMBL:X86798; NID:g806573; PIDN:CAA60493.1; PID:g806574

A:Accession: S56127

A:Molecule type: protein

A:Residues: 40-54 <MAX>

A:Genetics:

A:Gene: celE

C:Function: A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-33/Domain: signal sequence #status predicted <SIG>

F:40-570/Product: cellulase #status experimental <MAT>

F:380-417/Domain: glycosidase GCM domain homology <GCM>

F:472-569/Domain: bacterial cellulose-binding domain homology <BCB>

Query Match

Best Local Similarity 7.8%; Score 203.5; DB 2; Length 570;

Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

QY 14 SLTLASSILFVSGTSTANANGFTVSGTLYDANGPFWNRGIN-----HGHMYKD 65
 DB 22 AIIILGSGILGGVSAQADVAPLSVQGNKIL-ANGOPASESGSLFWSNTEWGEKYNA 80
 QY 66 QATTAIGTANTGANTVRIYLS--SDGGQWTKDD-THFRNLISLAEDNHLVAPEV 118
 DB 81 QVSWMLK--SDWNNKLVRAAGVEDESGYITDPAANDRYQVVD-AALANDMYIILDM 137
 QY 119 HDATGYDSIASLNRADVWTEKRSALIGKEDTYIINANEMFG-SWEGDAMADGYQA-1 176
 DB 138 HMAHQYOS-----QALAFQDEM-ARKYGANNHYEYIENPELOVWSMT--IKPYAQAVI 189
 QY 177 PRLNAGNHTLMTDAAGWGFPOSIDHYGREVFADQ--QNTMFSIHMEYVAGNAG 233
 DB 190 AAIRAIDPDNLIIVGTPTWSQ-----DVIDAANDPITGYQIVATTLHF--YAGTHGQ 239
 QY 234 QVRETRIDVTLNODLALVIGFGRHTNDVDEATIMSSEORGVGLAWSKNGKGPWEYL 293
 DB 240 YLRKAQALNRGIALVTEMGSYNANGD-----GAVANSE----- 275
 QY 294 IDLSNDMAGNNLTAMGNTYNGYGLRETSLSTVFTGGSDGTSPTLYLDFEGSMQGW 353
 DB 276 ---TNAWVSFMKT--NHISNANMALNDKVEGASALVPGASANG-----GW 315
 QY 354 TGSSSLSGGPNV-----VTEW-----SSKGSLSLKADIOSNSQHLHYQNTSLQONSRIQ 405
 DB 316 VNSQLTASGALAKSLISGMPSTNTSSSSSAVSSQTVSSSQ--APVYSSSSSTASSVVS 373
 QY 406 ATV--KHAWGSGVNGMARTLVYKIGGYT-----YGGSEVPYINGSGTTLISL 452
 DB 374 SAVSQGQCMW-----YGLIYPLCSTTTNGMENNASCARATCGOPAPMGIVGGSTSS- 428
 QY 453 DLSNVQNTSLQONSRIQATVKAHNGSVG--NGATPALLYKTGTGTYTSGSEFPIN 482
 DB 429 -----QASSSVYSSSSSIYSSSSSSSV 453

RESULT 13

A27198 cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain IF03034)

N:Alternate names: endo-1,4-beta-glucanase

C:Species: *Bacillus subtilis*

C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999

C:Accession: A27198

R:Nakamura, A.; Uozumi, T.; Bepko, T.
 Eur. J. Biochem. 164, 317-320, 1987.
 A>Title: Nucleotide sequence of a cellulase gene of *Bacillus subtilis*.
 A:Reference number: A27198; MUID:87190397
 A:Accession: A27198
 A:Molecule type: DNA
 A:Residues: 1-499 <NAR>
 A:Cross-references: GB:M28332; NID:g142670; PIDN:AA22307.1; PID:g142671
 A:Experimental source: strain IFO3034
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 8.5e-06;
 Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

```

QY 6 KRFSTTSLASSLTF-----VSGSTANANG-FVSGTTLTDANGNPFVNRG 55
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 KRSISFICLLITVLMGGLQSPASAGTKPAKNGLSIKGTOLVNRGKAVQLKG 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 I-NHGAWYKQDATTAEIGANTGANVRIYLSDGQWTKD-----IHT----- 99
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 ISSHGLQWTD-----FVNKDSLKMLRDMGIVTFRAAYTADGGYI 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 100 -----VRNLSLAEDNHLVAPEVDATGYDSIASINRAVDYIEKRSALIGKEDTV 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 104 DNPSYKMKKEAVEAKELGIYIIDWHLNDGNPNQNEKAKFEKEM-SSLYGNTPNV 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 IINIANEMFG--SWEGD--AMADGYKQALPRLNAGNLNTLWDAAGCQFPOSIDYGR 207
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 163 IYEIANEPGDVNMKRDIPYAE--EVSIVIRKNDPDITIV--GIGTWSQDVND--- 213
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 208 EYFNADPQ---RNTMSIMTYEAGNAGSOVNTNDRVNLQDLALVIGERHRTNGD--- 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 214 ---AADQDKANVMYALHF--YAGHGOSLDRKANVYALSKAPFVTEWGSADASNGG 268
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 263 ---VDEATIMSYEORGVGLWAMSKNGPEWY--EYLDLSN-DWAGNANLWAGNTIYNG 318
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 VFLDQSR-----EWANLTDKSKISWVNMNL----- 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 319 LRETSLSLSTVETGGSGDGTSPITLYDFEGSMQGTSSLSGGPMAVTEWSSKSHSLKA 378
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 -----SDKQESSALK-----PGASKTG-WEPLTDLTASGTF-VRE 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 379 DI--QLSSNSQHYLHYQNTSLQNSRIQATYKHAMGSGVGNMTRALVYKIGCHGT--- 433
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 328 NILGNKDSIKKEPPEPAQNPQAEENG-ISOYKAGAGGAGVNSQIRPOLIKKNGNATVDL 386
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 434 -----WYSGSEFVPIGSSGTTLSLDSNV 457
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 387 KDVTARYWYNAR-----NKGQNECDYAKOI 411
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14
 A26874
 cellulase (EC 3.2.1.4) precursor - *Bacillus subtilis* (strain DLG)
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
 C:Accession: A26874; B26874
 R:Robson, L.M.; Chambliss, G.H.
 J. Bacteriol. 169, 2017-2025, 1987
 A>Title: Endo-beta-1,4-glucanase gene of *Bacillus subtilis* DLG.
 A:Reference number: A26874; MUID:87194581
 A:Accession: A26874
 A:Molecule type: DNA
 A:Residues: 1-508 <ROI1>
 A:Cross-references: GB:M16185; NID:g143007; PIDN:AA22496.1; PID:g143008
 A:Experimental source: strain DLG

A:Accession: B26874
 A:Molecule type: protein
 A:Residues: 39-53 <ROI2>
 A:Experimental source: strain DLG
 A:Comment: The authors believe Met-1 and Met-2 may be alternate initiators
 C:Comment: The low molecular weight of the mature protein suggests carboxy-terminus
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: alternate initiators; extracellular protein; glycosidase; hydrolase;
 F:1-38/Domain: (or 2-38) signal sequence #status predicted <SIG>

Query Match 7.2%; Score 188.5; DB 2; Length 508;
 Best Local Similarity 21.1%; Pred. No. 1.8e-05;
 Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17;

```

QY 6 KRFSTTSLASSLTF-----VSGSTANANG-FVSGTTLTDANGNPFVNRG 55
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 KRSISFICLLITVLMGGLPSPASAGTKPAKNGLSIKGTOLVNRGKAVQLKG 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 I-NHGAWYKQDATTAEIGANTGANVRIYLSDGQWTKD-----IHT----- 99
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 71 ISSHGLQWTD-----FVNKDSLKMLRDMGIVTFRAAYTADGGYI 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 100 -----VRNLSLAEDNHLVAPEVDATGYDSIASINRAVDYIEKRSALIGKEDTV 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 113 DNPSYKMKKEAVEAKELGIYIIDWHLNDGNPNQNEKAKFEKEM-SSLYGNTPNV 171
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 IINIANEMFG--SWEGD--AMADGYKQALPRLNAGNLNTLWDAAGCQFPOSIDYGR 207
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 172 IYEIANEPGDVNMKRDIPYAE--EVSIVIRKNDPDITIV--GIGTWSQDVND--- 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 208 EYFNADPQ---RNTMSIMTYEAGNAGSOVNTNDRVNLQDLALVIGERHRTNGD--- 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 223 ---AADQDKANVMYALHF--YAGHGOSLDRKANVYALSKAPFVTEWGSADASNGG 277
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 263 ---VDEA-TIMSYEORGVGLWAMSKNGPEWY--EYLDLSN-DWAGNANLWAGNTIYNG 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 278 VFLDQSRWNLNLDKSNISWVNMNLSDKQESSALKPGASKTGWPLDITLTSAGTFVREN 337
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 PYGLRETSR 324
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 338 INGRDSTR 346
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 15
 G69593
 cellulase (EC 3.2.1.4) bgIC precursor - *Bacillus subtilis*
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: G69593; A26114; I40353; S49103; I39803
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Enlian, K.D.; Erington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lartino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadleir, J.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schwoerer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akuchil, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yato, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A:Reference number: A69580; MUID:98044033
 A:Accession: G69593
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-508 <KUN>

A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CA813696.1; PID:e118347.1;
 A:Experimental source: strain 108
 R:MacKay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moravelli, F.; Sel
 Nucleic Acids Res. 14, 9159-9170, 1986
 A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
 A:Reference number: A26114; MUID:87066783
 A:Accession: A26114
 A:Molecule type: DNA
 A:Residues: 10-508 <MAC>
 A:Experimental source: strain PAP115
 A:Note: Part of this sequence, including the amino end of the mature form, was confirmed
 R:Landahl, V.; Aa, K.; Tronsmo, A.
 Antonie Van Leeuwenhoek 66, 327-332, 1994
 A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis C
 A:Reference number: I40353; MUID:95225656
 A:Accession: I40353
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-291, 'N', 293-508 <LIN1>
 R:Landahl, V.; Aa, K. K. EMBL Data Library, June 1992
 A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
 Submitted to the EMBL Data Library, June 1992
 A:Reference number: S24239
 A:Accession: S24239
 A:Molecule type: DNA
 A:Residues: 10-291, 'N', 293-508 <LIN2>
 A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
 A:Experimental source: strin CK-2
 R:Wolf, M.; Geetzl, A.; Borriess, R.
 submitted to the EMBL Data Library, December 1993
 A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis: cons
 A:Reference number: S49103
 A:Accession: S49103
 A:Molecule type: DNA
 A:Residues: 10-508 <MOI>
 A:Cross-references: EMBL:Z29076; NID:g509266; PIDN:CA82317.1; PID:g509267
 R:Seo, Y.S.; Lee, Y.H.; Park, U.H.; Kang, H.
 Korean J. Microbiol. 24, 236-242, 1986
 A:Title: Analysis on the nucleotide sequence of the signal region of Bacillus subtilis
 A:Reference number: I36003
 A:Accession: I36003
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 10-13, 'V', 15, 'N', 17-18, 'V', 20-21, 'P', 23, 'A', 25-26, 'AI', 29-31, 'P', 33, 'PQ', 36-
 A:Cross-references: GB:M8634; NID:g142657; PIDN:AAA22300.1; PID:g142658
 A:Experimental source: strain ATCC 6633
 C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal pro
 C:Genetics:
 A:Gene: bglc
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
 C:Keywords: cellulose degradation
 C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:39-508/Product: cellulase #status predicted <MAT>

Query Match 6.9%; Score 181; DB 2; Length 508;
 Best Local Similarity 19.3%; Pred. NO. 5.8e-05;
 Matches 100; Conservative 102; Mismatches 186; Indels 128; Gaps 26;

QY 6 KRFTSLTLLASTL-----VSGTSTANNSG-FYVSGTTLTDANGNPFVARG 55
 DB 11 KNSISFTICLLTLLTMGMATSPASAGTTPVAKNGQLSINGQLVNRDGAVALQKG 70
 QY 56 I-NHGAWYKDDAT-TAEGIANGTANTV--RIVISDGGQNTKDDI-HTVANLISLAE 109
 DB 71 ISHGLQMTGEYVNRKSLKLRDMDGIFVRAMYTADGTYIDNPSYKKNKVEAVEAFAE 130
 QY 110 NHELVANPEVHADAGYSIASLNRAVDYWTENRSGALGKEDTVIINANEMFG--SWEGD- 166
 DB 131 LGIYIIDWHIINDGNPNONKREKAEFEKEM-SSLTGNTPNYIYEIANEPNDGVNKKRDI 189

QY 167 -ANADGYKQALPRLNAGLNTLTAMDAGWGQFPOSIDYGEVFNADPO---RTIMESI 222
 DB 190 KPYAE---EIVSIRKNDNDNIIIV---GTGTWSDVDN-----AADDQKDNAMVMTAL 237
 QY 223 HMYEYAGNNAQVRTNIDVNLNODLALVYGEFGHRTNGD---VDEA-TIMSIEQRCV 277
 DB 238 HF--YAGTHQFLADKANALSKGAPITFTEWGTSDASGNGGVFLDQSRKMLKIDSXTI 295
 QY 278 GWTAMSKNGKPEWEYL---DISNDWAGNNLTANGNTIVNPGYGLRERS----- 324
 DB 296 SWVMNINSDKQESSSALKPKPASKRTGWRSLDSASTFRENILGDKSTKIDIPETPSKD 355
 QY 325 -----LSYFTLGGSDGDSPTTLTYDEGSMQGTGSSLSGGPWAATWESSKGSLSK 377
 DB 356 KPYQENGISVQYRAG-----DGSN-----NSNQIR 380
 QY 378 ADIQLSN-----SGHYLHVIONTSLOQNSHIOATVKAHNGSVGNMTARLY- 426
 DB 381 PQLQIKKNGNTYVDLKVYARWYKAKNKG--QNF-----DCDYAQIGCGNVTAKFVT 431
 QY 427 --KTGHGTYW-----SGSFVPIINGSSGTTLSLDSN 456
 DB 432 LHKPKQAGADTYLLELGFKNKGLAP--GASTGNITQLRLHN 467

Search completed: December 19, 2000, 16:10:59
 Job time: 640 sec

This Page Blank (usp10)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:29:24 ; Search time 131.22 Seconds
(without alignments)
119.300 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNRFRKIRITSLILASS.....QSASDSSGOTSTIDNVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 737.5 | 28.3 | 383 | 1 | MANA_STRLI |
| 2 | 638 | 24.5 | 1331 | 1 | MANB_CALSA |
| 3 | 427 | 16.4 | 513 | 1 | MANB_BACSM |
| 4 | 234 | 9.0 | 448 | 1 | GUN_CLOAB |
| 5 | 229.5 | 8.8 | 400 | 1 | GUNS_BACAG |
| 6 | 226.5 | 8.7 | 488 | 1 | GUN1_BACSA |
| 7 | 226 | 8.7 | 409 | 1 | GUN2_BACSA |
| 8 | 220.5 | 8.5 | 505 | 1 | GUNV_ERMCA |
| 9 | 219.5 | 8.4 | 444 | 1 | GUNV_ERMCA |
| 10 | 207 | 7.9 | 504 | 1 | GUNW_ERMCA |
| 11 | 188.5 | 7.2 | 499 | 1 | GUN1_BACSU |
| 12 | 181 | 6.9 | 499 | 1 | GUN2_BACSU |
| 13 | 180 | 6.9 | 466 | 1 | GUNS_THERU |
| 14 | 171 | 6.6 | 499 | 1 | GUN3_BACSU |
| 15 | 170.5 | 6.5 | 584 | 1 | GUN3_CLOAE |
| 16 | 168 | 6.4 | 800 | 1 | GUN_BACSI |
| 17 | 166 | 6.4 | 941 | 1 | GUNA_BACSE |
| 18 | 165.5 | 6.3 | 459 | 1 | GUNA_STRLI |
| 19 | 164.5 | 6.3 | 825 | 1 | GUN3_BACSA |
| 20 | 160.5 | 6.2 | 510 | 1 | YXPD_SCHRO |
| 21 | 155 | 5.9 | 814 | 1 | GUNE_CLOFM |
| 22 | 151.5 | 5.8 | 562 | 1 | GUNA_ACICE |
| 23 | 147.5 | 5.7 | 429 | 1 | GUNA_BUTFI |
| 24 | 146.5 | 5.6 | 747 | 1 | GUND_CELFI |
| 25 | 141 | 5.4 | 312 | 1 | GUNA_RUMAL |
| 26 | 140.5 | 5.4 | 426 | 1 | GUN4_BURSO |
| 27 | 133 | 5.1 | 1039 | 1 | AC43_BCOLI |
| 28 | 132.5 | 5.1 | 1608 | 1 | HYLA_SERVA |
| 29 | 132 | 5.1 | 438 | 1 | EXG_CANAL |
| 30 | 131 | 5.0 | 515 | 1 | GUND_CLOCL |
| 31 | 131 | 5.0 | 517 | 1 | GUNA_CLOLO |
| 32 | 129 | 4.9 | 566 | 1 | GUNG_CLOFM |
| 33 | 129 | 4.9 | 2003 | 1 | YDBA_BCOLI |

| | | | | | |
|----|-------|-----|------|---|------------|
| 34 | 128.5 | 4.9 | 890 | 1 | BCN5_CLOPE |
| 35 | 127 | 4.9 | 426 | 1 | GUNZ_ERMCH |
| 36 | 123.5 | 4.7 | 645 | 1 | LIP1_PHOUD |
| 37 | 123.5 | 4.7 | 711 | 1 | CDGT_BACST |
| 38 | 122.5 | 4.7 | 1039 | 1 | GUNB_CALSA |
| 39 | 118.5 | 4.5 | 1829 | 1 | FRPC_NEIMC |
| 40 | 116.5 | 4.5 | 440 | 1 | GUNB_CLOCL |
| 41 | 116.5 | 4.5 | 976 | 1 | PIBP_ADEB3 |
| 42 | 115 | 4.4 | 1656 | 1 | OMP_RICUA |
| 43 | 114.5 | 4.4 | 865 | 1 | HYRE_ECOLI |
| 44 | 114 | 4.4 | 802 | 1 | PEFC_SALTY |
| 45 | 114 | 4.4 | 2334 | 1 | WAPA_BACSU |

ALIGNMENTS

| RESULT | ID | MANA_STRLI | STANDARD | PRT | 383 AA. |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|----------|-----|---------|
| AC | P51529 | | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | | |
| DE | MANNA ENDO-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-MANNANASE) (1,4-BETA-D-MANNAN MANNANOHYDROLASE). | | | | |
| GN | MANNA. | | | | |
| OS | Streptomyces lividans. | | | | |
| CC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | | | |
| CC | Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 36-42. | | | | |
| RC | STRAIN-66 / 1326; | | | | |
| RA | MEDLINE: 93207541. | | | | |
| RT | Arcand N., Kuepfer D., Paradis F.W., Morosoli R., Shareck F.; "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence of the manA gene and characterization of the enzyme." | | | | |
| RL | Biochem. J. 290:857-863(1993). | | | | |
| RN | [2] | | | | |
| RP | REVISIONS TO C-TERMINUS. | | | | |
| RC | STRAIN-66 / 1326; | | | | |
| RA | Shareck F.; | | | | |
| RL | Submitted (APR-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES CELSIUS AND A PH OF 6.8. | | | | |
| CC | -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS. | | | | |
| CC | -1- SUBUNIT: MONOMER. | | | | |
| CC | -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES). | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL: M92297; AAA26710.2; -. | | | | |
| DR | INTERPRO: IPR001547; -. | | | | |
| DR | PFAM: PF00150; cellulase; 1. | | | | |
| DR | PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1. | | | | |
| FW | Hydrolase; Glycosidase; Signal. | | | | |
| FT | SIGNAL | | | | |
| FT | CHAIN | | | | |
| FT | DOMAIN | | | | |
| SEQUENCE | 383 AA; 39681 MM; SDB4B407C64E9AC3 CRC64; | | | | |

Query Match 28.3%; Score 737.5; DB 1; Length 383;
Best Local Similarity 44.4%; Pred. No. 1.8e-43;

Matches 160: Conservative 56; Mismatches 121; Indels 23; Gaps 9;

OY 13 LSLASSLLEPVSSTANAN---GFYSGTTLTDANGNPPVAGINHGAWYDQATT 69
DB 19 LGLFA-----IAGSAGAAEAAAGIHNSGRVYEGNSAFVAGVNAHYWYDR-TG 72
OY 70 ATEGIANTGANTVRIYLSGQWGTNDKHITVNLISLAEDNHLVAVPEHATGY--DSI 127
DB 73 SIADIAAGANTVRYVLSGGGRWTTSSAEVSALIGOCKANKVICYLEVHDTTGKDGGA 132
OY 128 ASLNDVAVDIEMRSAL-IGKEDVYIINIANWFSGWEGDAMADYKQALPRLNAGINH 186
DB 133 TSLDAGDYVWGVKSAWRAQEDYVYVNGINEPNTNAAWTDATKSAIGLRGAGLGH 192
OY 187 TLMNDAAGMC-FPOSIDYGREVNADPQRTNFSIHMEYAGNASVYRNIVRLNQ 245
DB 193 ALMDVAPMNGDMSGTMSNAVSAPSDRTVPSIHNG-VYDTAEVRDYLNAFNGN 251
OY 246 DLATVIGFGRHRTNGDVDEATIMSSEORVGLAMSKWNGPMEYTLDSNDWAGNLL 305
DB 252 GLPIYVGEFGDSDGNDDEDAIMATAGSLGYLGKMSNGGVEYIDMNGDPNLSL 311
OY 306 TAWGNTVNGPYGLRETSRLSTVF-TGGSGDGTSTP-----TLVDEGSMQWGTGS 356
DB 312 TSMGNRIYSGNGIATSRATATVYGGGGSTGTAPNGPYCVNGGASDPDGDGMEWENS 371

RESULT 2
MANB_CALSA STANDARD; PRG: 1331 AA.

AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BETA-MANNANASE/ENDOGALACTANASE A PRECURSOR (INCLUDES: MANNAN ENDO-1,4-
DE BETA-MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-
DE MANNAANASE); ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULOSE)).
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermomonadobacter group; Caldicellulosiruptor.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 93119139.
RA Glibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
RT multidomain enzyme";
RL Appl. Environ. Microbiol. 58:3864-3867 (1992).
RN [2]
RX MEDLINE: 91247819.
RA Luthi E., Jussant N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in *Escherichia coli* of a
RT gene coding for a beta-mannanase from the extremely thermophilic
RT bacterium 'Caldocellum saccharolyticum'";
RL Appl. Environ. Microbiol. 57:694-700 (1991).
CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
CC GALACTOGLUCOMANNANS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
CC CELSIUS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L01257; AAA71887.1; -;
DR EMBL: M30603; AAA72861.1; -;
DR PIR: B43745; B43745.
DR PIR: A48954; A48954.
DR HSSP: Q06851; INBC.
DR INTERPRO: IPR001547; -;
DR INTERPRO: IPR001956; -;
DR PFAM: PF00942; CBD_3; 2.
DR PFAM: PF00150; cellulase, 1.
DR PROSITE: PS00659; GLYCOSYL-HYDROL-F5; 1.
DR HydroLase: Glycosylase; Cellulose degradation; signal;
KW Multifunctional enzyme.
FT SIGNAL 1 41
FT CHAIN 42 1331
FT DOMAIN 42 325
FT DOMAIN 326 361
FT DOMAIN 362 518
FT DOMAIN 519 564
FT DOMAIN 565 720
FT DOMAIN 721 780
FT DOMAIN 781 1331
FT ACT_SITE 162 162
FT ACT_SITE 257 257
FT ACT_SITE 338 338
FT CONFLICT 340 346
FT SEQUENCE 1331 AA; 146892 MW; FFECA51BB8D870E0 CRC64;

Query Match 24.5%; Score 638; DB 1; Length 1331;
Best Local Similarity 31.4%; Pred. No. 5.9e-36;
Matches 166; Conservative 84; Mismatches 208; Indels 70; Gaps 14;

OY 6 KRISITSLASSILFVS-----GISTANNSGFYSGTTLTDANGNPPVAGIN 57
DB 9 KRWLSVLTCTVFLNIFLFIANTVILPRVGAATSN-----DGVKKIDTS-----TLIGTN 57
OY 58 HGHWYKDOATTAIEGNTGANTVRIYLSGQWGTNDKHITVNLISLAEDNHLVA-VP 116
DB 58 HAHCHYRRLTLALRGISWGNVSRYVLSNGYKWKIKPASEVANIISLSLGFKALITL 117
OY 117 EYHDATGY--DSIASLRAVDYIEMRSALIGKEDVYIINIANWFSGWEGDAMADGYK 173
DB 118 EYHDPTGYGEGDAGCSLAQAVEYKEIKSVLDGNEDEFITINIGNEPQNNYQMMVNDTK 177
OY 174 QALPRLNAGINHTLMDADAGMC-FPOSIDYGREVNADPQRTNFSIHMEYAGGNA 232
DB 178 NAIKRLNAGKFKHITMDAPMGQDWSMTMDNMSIMEDPLNVLVYSIMYG-VYNTA 236
OY 233 SOVFRNIDRVNODIATVIGFGRHRTNGDVDEATIMSSEORVGLAMSKWNGPMEWE 292
DB 237 SKVEEYIKSPDKPLPLYIGFEGHDTGDDPEDAIYRAQYIIGLFSWCGNSSTYVG 296
OY 293 YLDSNDWAGNNTLAWGNTVNGPYGLRETSRLSTVFVTGGSDGT-SPITLYDEGSMQ 351
DB 297 YLDMVNMNDPNPPTPMGQWYKTAIGTSSPTPTSTVPTPTPTPTPTATPTPTPTPT 356
OY 352 GWTGSSLSG-----GPAVITWSSKSGSLADYQLSNSQHYLHYT 393
DB 357 PVSPATPSGQIKVLYANKETSTNTTIRPWL-----KVNSGSSSIDLSFTVTRRYWTV 410
OY 394 QNTSLQNSRQIATVYKHNAGSVNGWTAFLYTGK-----GYWWSGSFVPI 442
DB 411 -----DGRQSAI--SDMAQICASVNTFFVYLVSSVSAGADYLLIEIGFSGAGQLPG 462
OY 443 NGSSGTTLSLDLSNVONLSQVREIG-VQFQASDSSGQTSYIDNVIV 489
DB 463 KDTGEIQRFKDKDMSVNGQNDMSWISQMSYENKRYATIDGVIV 510


```

CC      -1- SUBUNIT: MONOMER.
CC      -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC              HYDROLASES).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF067428; AACJ9169.1; -.
DR      PDB; 1A3H; 16-MAR-99.
DR      PDB; 2A3H; 16-MAR-99.
DR      PDB; 3A3H; 16-MAR-99.
DR      INTERPRO; IPR001547; -.
DR      Pfam; PF00150; cellulase; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW      Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT      ACT_SITE 165..165
FT      ACT_SITE 254..254
FT      ACT_SITE 400 AA; 44702 MW; 3Pc66FB9BC36FEF CRC64;
SQ      NUCLEOPHILE.
SQ      SEQUENCE
SQ      Query Match 8.8%; Score 229.5; DB 1; Length 400;
SQ      Best Local Similarity 23.4%; Pred. No. 9.56-09;
SQ      Matches 97; Conservative 64; Mismatches 172; Indels 81; Gaps 19
SQ
QY      6 KRISITSLILLASIIIFVSGSTANANS-----GFVSGFTLLIDANGNPFVNRGI-NH 58
QY      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      2 KRIITFIIVALLVALFSIGNTIADNDSDVEEHOLSISSNGELYNENGEQVLKGMSSH 61
QY      59 GHAWT-----KDQAATTAIEGIANGTANTVRIVL--SPGGGWTKDDI-HYVRNL 103

```

```

0Y      104 ISLADNHLVAVPEVDATGDSIASLANVAQVMTEMASALIGKDDPIIANEMFGSW    163
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
DB      112 YEAALDIDLYIIMHLLSDNDPNIKYEKKDPFDEM-SELGYDPNVITYELANEPNGS-   169
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
OY      164 EGDAAVAGX-----QAIPRLRNAGLNHTLIWPAAGAOGPPOSINDHYGREVNADPQRNTM   219
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
DB      170 -DYWGNGRIKRYAEVEVPIIRNMDPNIIIV--GTGTSGDVH-HAADNLADP--NYM     222
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
OY      220 FSIHMYETAGGNASOVRTNDRVLNODLALVTGEGRHRTNGD----YDEATI-MSYSEQ     274
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
DB      223 YAFHR--YAGTHGONLRDOVDYALLDGGALFESEMGTSATDGCGVFIDEAQWIDPFDE     280
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
OY      275 RAYGLTAYS-----WKSGCPMEWEYLDSNMAGNNLTJANGNTIVNGPYGLARETS       323
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
DB      281 KULSMANNSLHKDESSAALMPGANP-----TGWTLEALSPTGFVRE---KIRESA       330
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
OY      324 RLSTVFTGGSGDCGSTPFTLYDFEGSNQGTGSSL-----SGGPNAVITWS         369
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :
DB      331 SIIPSPDTPPSPPGEPDPFTPSPDGEPAPMDPOIYTNEIYTHNGOLMOAKMT          384
        : - : - : - : - : - : - : - : - : - : - : - : - : - : - : - :

RESULT      6
GUN1_BACS4      : STANDARD;                PRT;           488 AA.
AC             P06566;
DT              01-JAN-1988 (Rel. 06, Created)
DT              01-JAN-1988 (Rel. 06, Last sequence update)
DT              15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLOUCANASE A,(EC 3.2.1.4) (ENDO-1,4-BETA-GLOUCANASE) (CELLULOSE).
GN CELA..
OC Bacillus sp. (strain N-4)..
OS Bacteria; Firmicutes; Bacillus/Clostridium group;.
CC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87056924.
```


GUNV_ERWCA ID GUNV_ERWCA STANDARD: PRT: 505 AA.

AC 047096: 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE ENOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)

DE (CELLULOSE V).

GN CELV.

OS Erylnia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SCRI193;

RX MEDLINE: 94067016.

RA Cooper V.J.C., Salmond G.P.C.;

RT "Molecular analysis of the major cellulase (Celv) of Erylnia carotovora: evidence for an evolutionary 'mix-and-match' of enzyme domains."

RL Mo., Gen. Genet. 241:341-350(1993).

CC -1- FUNCTION: ENOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES CELSIUS.

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: X76000; CAA53592.1; -.

DR HSSP: Q06851; INBC.

DR INTERPRO: IPR001547; -.

DR INTERPRO: IPR001956; -.

DR PFAM: PF00942; CBD_3; 1.

DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

KM Cellulose degradation; Hydrolase; signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 505 ENDOGLUCANASE V.

FT DOMAIN 32 334 CATALYTIC.

FT DOMAIN 333 352 LINKER.

FT DOMAIN 333 505 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SEQUENCE 505 AA; 54900 MW; DBBA9337BB4D2623 CRC64;

Query Match 8.5%; Score 220.5; DB 1; Length 505; Best Local Similarity 21.8%; Pred. No. 5.3e-08; Matches 116; Conservative 69; Mismatches 209; Indels 137; Gaps 21;

QY 2 NNGFKIFSTISLASSLFSVSGSTANANG-YYVSGTLYDANGNPFVARGI-NHG 59

DB 6 NQIVRLITGVYTVVIGMSISFALSATVEYRGQISIEGRVLVDQGRVQLRGISSHG 65

QY 60 HMYKQDATTAEIGIANTGANTVRIYVSDGGQWTKD----- 96

DB 66 LQFQGD-----YVKNDSMKVLRODQWGINVRVAMVTAADGYISNPSL 107

QY 97 IHTVRNLISLAEDNHLVAVPEVDATGYDSIASLNAADVIMKRSALLGKETVLIINIA 156

DB 108 ANKVEAVAAOSLGYIITIDHILSDNDPNIRYKAQKTFEFAM-AGLGSSPNVITIELA 166

QY 157 NEMFG--SMEGD--AWADGYKQAIPLRLNAGLHNTLAVDAAGQFPQSIHDYGREVFNA 212

DB 167 NEPNGVATWNGQIRPYA---LEVTDTIRSKOPDLITV---GIGTWSQDTHD---AADNQ 217

QY 213 DPORTMFSLIMEYAGNAGNSQVNTNDRVLDLALVIGFGRHTNGD-----VDEAT 267

DB 218 LPDNTMYALHF--YACTHQGLFDRIDYAGSRGALFVSWMGSDASGPPLPESQT 275

QY 268 IMYSSEQRGVWLMASKNGKNGPEWEYL-----DLSNDNAGNLTAMGTIYNGPYGLAETS 323

DB 276 WIDFLNNGVSWVMWSLTDKSEASALAPGASKSGWTEONLSTGKFXV-----REQI 328

QY 324 RLSTVFTGGSGDGTSPPTLIDFEGSMQGTSSLSGPNVAVTWSKSGSHLRADILS 383

DB 329 R-----AGANLGGDITPTTPEPTNGNGITG-----DYYL- 359

QY 384 SNSQHYLVHONTLSQNSRIQATVKRANAGSVG--NGMTARLYKTHGYWYSGSFV 440

DB 360 ----QYRVNDNNS---DDAIRMAVNIKNGSTPIKLSDLQVRYFHD----- 400

QY 441 PINSSGTTLSDLSNQNLSQVREIGVQPSAD-----SSGQTSI 482

DB 401 -DGRPGANLPEVDANVGPNNIVTGTGPAASTKANRYLVTFSSGAGSL 449

RESULT 9

GUNV_ERWCA ID GUNV_ERWCA STANDARD: PRT: 444 AA.

AC Q59394; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE ENOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N)

DE (CELLULOSE N).

GN CELN.

OS Erylnia carotovora.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATROSEPTICA FCBR C18;

RX MEDLINE: 98299944.

RA Olesen O., Thomsen K.K., Weber J., Duns J.O., Svendsen I., Wegener C., von Wettstein D.;

RT "Transplanting two unique beta-glucanase catalytic activities into one multienzyme, which forms glucose."

RL Biotechnology 14:71-76(1996).

CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: L39788; AAC37033.1; -.

DR INTERPRO: IPR001547; -.

DR INTERPRO: IPR001956; -.

DR PFAM: PF00942; CBD_3; 1.

DR PFAM: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

KM Cellulose degradation; Hydrolase; signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 444 ENDOGLUCANASE N.

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

GN BGIC OR GLD.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-DLG;
 RX MEDLINE; 87194581.
 RA Robson L.M., Chambliss G.H.;
 RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG";
 RL J. Bacteriol. 169:2017-2025(1987).
 CC -1- CATALYTIC ACTIVITY: ENDODIHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M16185; AAA22496.1; ALT_INIT.
 DR PIR; A26874; A26874.
 DR HSSP; 006851; INBC.
 DR INTERPRO; IPR001547; -
 DR INTERPRO; IPR001956; -
 DR PFAM; PF00942; CBD_3; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 FT SEQUENCE 499 AA; 55187 MW; 339D04EE95A65BE1 CRC64;
 SQ
 Query Match 7.2%; Score 188.5; DB 1; Length 499;
 Best local Similarity 21.1%; Pred. No. 8e-06;
 Matches 78; Conservative 75; Mismatches 133; Indels 83; Gaps 17;
 QY 6 KRIFSTLSLLASSILF-----VSGTANANG-FTYSGTLYTDANGNPPVARG 55
 DB 2 KRISIFITCLIAVLTMGILPSPASAGTKPVAKNGOLSIKQTVNRDGAVALKG 61
 QY 56 I-NHGAWKQDAITALEGIANTGANTVRIYLSDDGQWTRKD-----IHT----- 99
 DB 62 ISHGLQWGD-----FVKNDSLAKWLRDMGIVFRAANTADGVI 103
 QY 100 -----VNLISLADENHLVAVPEVHDATGYDSIASINRAVDYIEMRSALIGEDTV 151
 DB 104 DNPVANKVKEAVEAKKEGIYIIDWHILNDGNPNONKEKAEFEKEM-SDIYGTPTV 162
 QY 152 ITIANEMFG--SWEGD--AMADGYQAIPLRNAGLNIHTAMDAGNQOFPQSHDYDR 207
 DB 163 IYEIANEPNDVWKKRIKPYAE--EVSIVIRKNDPDNIIT--GTGWSQDVND--- 213
 QY 208 EVFNADPQ--RMTFESIHYEYAGNAGVARTINIDRVLNODLALYIGEGHRTNGD-- 262
 DB 214 ---AADDQLKDAVNTALH--YAGHGSGLRKANALSKGAPIVTEGTGSDAGNGG 268
 QY 263 --VDEA-TINSYSEQRGVGLANSWKGNGPEWEYL---DLSNDWAGNNITANGNTIVNG 315
 DB 269 VPLDQSEMLNLYDSKNIWVNNWNLSDKQSSALKPGASKGTGWPDLTILASTGVREN 328
 QY 316 PYGIARESR 324
 DB 329 IRGTRDKSTK 337

RESULT 12
 GUN2_BACSU STANDARD; PRT; 499 AA.
 AC P10475;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULASE) (CMCase) (CELLULOSE).
 GN BGIC OR GLD OR EGLS.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-PAP115;
 RX MEDLINE; 87066783.
 RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
 RA Moravell F., Seligy V.;
 RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene";
 RL Nucleic Acids Res. 14:9159-9170(1986).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CK-2;
 RX MEDLINE; 95225656.
 RA Lindahl V., Ae K., Tronsmo A.;
 RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
 RT subtilis CK-2";
 RL Antoine Van Leeuwenhoek 66:327-332(1994).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Rose M., Entian K.;
 RT submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE OF 30-45.
 RC STRAIN-CK-2;
 RX MEDLINE; 95225655.
 RA Ae K., Flensburg R., Lindahl V., Tronsmo A.;
 RT "Characterization of production and enzyme properties of an
 RT endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
 RT compost soil";
 RL Antoine Van Leeuwenhoek 66:319-326(1994).
 CC -1- CATALYTIC ACTIVITY: ENDODIHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z29076; CAA82317.1; -
 DR EMBL; X04689; CAA28392.1; -
 DR EMBL; X67044; CAA47429.1; -
 DR EMBL; Z73234; CAA97610.1; ALT_INIT.
 DR EMBL; Z99113; CAB13696.1; ALT_INIT.
 DR PIR; A26114; A26114.
 DR HSSP; 006851; INBC.
 DR SUBTILIST; BG10437; BGIC.
 DR INTERPRO; IPR001547; -
 DR INTERPRO; IPR001956; -
 DR PFAM; PF00942; CBD_3; 1.
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 499
 FT ENDOGLUCANASE.


```

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCase) (CELLULOSE).
GN BGIC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BSE616;
RX MEDLINE; 91299280.
RA Park S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616."
RL Agric. Biol. Chem. 55:441-448(1991).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01057; BAA00859.1; -.
DR PIR; J0111; JN0111.
DR HSSP; Q06851; INBC.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR001956; -.
DR PFM; PF00942; CBD_3; 1.
DR PFM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; signal.
KW SIGNAL
FT CHAIN 1 29 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NOCLOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55169 MW; 2E821E3DBBACA04 CRC64;

```

```

Query Match 6.6%; Score 171; DB 1; Length 499;
Best Local Similarity 19.0%; Pred. No. 0.00013;
Matches 98; Conservative 104; Mismatches 189; Indels 126; Gaps 25;

6 KRIFSTILLASSILF-----VSGTSTANNSG-FYVSGTTLXDANGNPFVMSG 55
2 KRISSTIFCILLTLTGMGMLASPASAGCTKTPVAKNQLSKTGQVYNRGKAVQLKG 61
56 I-NHGAWTKDQAT-TALGCIANTGANTV--RIVLSDGGQWTKDI-HTVNLISLAD 109
62 ISSHGLWGTGEYVKNLSKLWLRDWDGIVFRAMYTADGIIIDNPVKKMKKAVAKRE 121
110 NHVYANPEVHDAGVDSISLNAVDYNTIEMRNALIGKEVYIINANEMFG--SNEG 166
122 LGIVYIIDHILNDGNPNONKEKAKEFKEM--SLLGNTPNVYIETLANPENGVDNKRDI 180
167 -AMADGYKOAIPRLNAGLNTLMDVDAAGVGPPOSIIHDYGRFVNADPO---RNTMFSI 222
181 KPAE---EYISVIRKNDPNITIV---GTGTSQDVND-----AADQLKANAMDL 228
223 HMYEYAGNAGSOVRTNIDRYLNDLALVIGFGRHTNG---VDEA-TIMSYSGRGV 277
229 HF--YAGTHGQFLDRKANALSLSGAPLIVTEMGTSASGSGVGFLLDQSRMAYLDSKI 286
278 GWLAWMKNGKGPWEYL-----DISNDWAGNLTAMGNTIYNGYGLRETSR----- 324
287 SWNNMNLSPKQESSALKPKGASRTGMRSLDLSASGTFVRENTIAGTKRDLPEPPAND 346
325 -----LSTVFRTGGSGDGTSPITLYIDFGSMQGWTSLSGGMNAVTEWSSKGSLSLK 377

```

```

DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE D)
DE (CELLULOSE D) (CECCD).
GN BGIC.
OS Bacillus/Clostridium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35319;
RX MEDLINE; 92009193.
RA Shima S., Igatahshi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RT celCD, of Clostridium cellulolyticum."
RL Gene 104:33-38(1991).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIOSYLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- PATHWAY: CELLULOSE DEGRADATION.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90341; BAA14354.1; -.
DR PIR; J01229; J01229.
DR HSSP; P17901; IEDG.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002048; -.
DR INTERPRO; IPR002105; -.
DR PFM; PF00404; DockertIn_1; 2.
DR PFM; PF00150; cellulase; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; signal.
FT SIGNAL 1 24
FT CHAIN 25 584 ENDOGLUCANASE D.
FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
FT DOMAIN 329 353 PRO/THR-RICH (LINER).

```

FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 530 584 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 530 552 1.
FT REPEAT 562 584 2.
FT ACT_SITE 159 159 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 264 264 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;

Query Match 6.5%; Score 170.5; DB 1; Length 584;

Best Local Similarity 20.5%; Pred. No. 0.00017;
Matches 119; Conservative 91; Mismatches 204; Indels 167; Gaps 30;

QY 6 KRIFITITLLASILFVSGSTANANGFVSGTLYDA-NGNPYMR---GINHG- 60
DB 2 KRIILALITSCSTINSFLPMS-----VYGAINSQDMVKMGIGMILNT 44
QY 61 -----AMYKQATTALEGIANGTANTVRIYLSDGQWTKDITVRNLISIAEDNHY 113
DB 45 FDAPFEGSMKAAQRYFDDFKQAGFKHRIPI---RM---DOHTLANSPTVDSNFLN 97
QY 114 AVPEY-----HDATGYDSIA-SLNRAVDYIEMRSALIGKEDVIINIA 156
DB 98 RIEFYIDMSISRGFVTYINSHDWTLMQNYSONIGRFKEIMEQIAQRFKSENLVFEIL 157
QY 157 NEMFGSMGEGDAMADQYKRAIPRLRNAGLHTLMVDAAGQFPOSIDHYGREVENADPQR 216
DB 158 NEPHGNITDSQJMDNKKRLNIRKTNPTRNVIIGAGYNSY-NSLSQI-EIPN-DP-- 211
QY 217 NTWFSIHMYE-YAGGNASQVNTNIDRVNLQDLAVIGFGRHTNGVDYEATINSYSQR 275
DB 212 NLIAFFHYIDPYSFTHQI-----GTWG---TKNDMD-AIAMVFNHVK 250
QY 276 GVGWLAMSMKNGPPEM-----EYLDLSNDMA--GNNLTAMGNTIVNGP 316
DB 251 -----KMSDKNNIPYILGEYGVMGSHDRISAQKMFDFVSDQAIHSFGSGAMD---NGV 301
QY 317 YGL-----RETSR-----LSVFTGGSGDGTSP-----TLX 344
DB 302 FGSVDNDMAFYNRDTRQPKETILNAILTTGTYDTPPTETNPDPPTAPATPAYGEOLIE 361
QY 345 DEEGSMQGTGSSLSGGFWAVTEWSSKSGSHIKADIDQLSSNSOHTLAVIONTSLOON--- 401
DB 362 DEEGAMQ-W--AAYSQVATASCKISSKSNNGLEITYAGSSNGYGVNDENHRQDMEX 418
QY 402 -SRIQATYKHAAMGWSVNGMTARLYKTGHGYTWYSGSFVPINGSSGTTLSLDSN--- 456
DB 419 WOKISFDIKSSNTNEVRLLIAEQSKIEGEDGBHW---TYVILKPTSWTTIEIPFSSFTKR 475
QY 457 -----VQNLQVREI-----GVQFOSASDSSGQTSIYIDNV 487
DB 476 MDYOPPADGSETFDLYKVGSLHFYATSNNSG--TLNIDNI 514

Search completed: December 19, 2000, 16:44:15
Job time: 891 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:00:19 ; Search time 363.46 Seconds
(without alignments)
175.883 Million cell updates/sec

Title: US-09-339-159-2

Perfect score: 2607
Sequence: 1 LNNGRKRTSITLSLLAASS.....QSASDSGGTSTIDNVIYE 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_14:*
2: sp.Archea:*
3: sp.Bacteria:*
4: sp.Fungi:*
5: sp.Human:*
6: sp.Invertebrate:*
7: sp.Mammal:*
8: sp.Ornithine:*
9: sp.Phage:*
10: sp.Plant:*
11: sp.Prodent:*
12: sp.Virus:*
13: sp.Vertebrate:*
14: sp.Unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|----------------------|
| 1 | 1622 | 62.2 | 516 | 2 066185 | 066185 bacillus cl |
| 2 | 814.5 | 31.2 | 395 | 2 069347 | 069347 vibrio sp. |
| 3 | 761.5 | 29.2 | 384 | 2 086599 | 086599 streptomyces |
| 4 | 634 | 24.3 | 279 | 2 092F13 | 092F13 thermomonas |
| 5 | 628 | 24.1 | 930 | 2 09RFX5 | 09RFX5 calidobacilli |
| 6 | 541 | 20.8 | 327 | 2 09RFX5 | 09RFX5 streptomyces |
| 7 | 253.5 | 9.7 | 1097 | 2 09ZAI7 | 09ZAI7 anaerobic t |
| 8 | 239.5 | 9.2 | 557 | 2 P94622 | P94622 clostridium |
| 9 | 228 | 8.7 | 389 | 2 059232 | 059232 bacillus sp |
| 10 | 203.5 | 7.8 | 570 | 2 059665 | 059665 pseudomonas |
| 11 | 193 | 7.4 | 499 | 2 045532 | 045532 bacillus su |
| 12 | 187.5 | 7.2 | 476 | 5 018453 | 018453 heterodera |
| 13 | 187.5 | 7.2 | 481 | 2 083012 | 083012 actinomyces |
| 14 | 186.5 | 7.2 | 387 | 2 031029 | 031029 erwinia car |
| 15 | 185.5 | 7.1 | 930 | 2 059290 | 059290 clostridium |
| 16 | 180 | 6.9 | 486 | 2 045430 | 045430 bacillus sp |
| 17 | 180 | 6.9 | 635 | 2 066065 | 066065 fibrobacter |
| 18 | 179 | 6.9 | 783 | 2 045554 | 045554 bacillus sp |

| | | | | | |
|----|-------|-----|------|-----------|---------------------|
| 20 | 178 | 6.8 | 749 | 2 059154 | 059154 anaerocellu |
| 21 | 177 | 6.8 | 499 | 2 052731 | 052731 bacillus sp |
| 22 | 173 | 6.6 | 621 | 2 007653 | 007653 cellvibrio |
| 23 | 169 | 6.5 | 821 | 2 059241 | 059241 bacillus sp |
| 24 | 161 | 6.2 | 478 | 5 016028 | 016028 globodera r |
| 25 | 160 | 6.1 | 278 | 2 09RFX5 | 09RFX5 streptomyces |
| 26 | 152 | 5.8 | 506 | 5 09UA57 | 09UA57 meloidogyne |
| 27 | 148.5 | 5.7 | 470 | 5 09U6M5 | 09U6M5 globodera t |
| 28 | 145.5 | 5.6 | 395 | 5 09U6M4 | 09U6M4 globodera t |
| 29 | 140.5 | 5.4 | 1335 | 12 089353 | 089353 parametrium |
| 30 | 138.5 | 5.3 | 910 | 3 P87211 | P87211 orpinomyces |
| 31 | 136.5 | 5.2 | 494 | 2 086099 | 086099 alteromonas |
| 32 | 136 | 5.2 | 438 | 3 09URL8 | 09URL8 candida alb |
| 33 | 133.5 | 5.1 | 319 | 5 018454 | 018454 heterodera |
| 34 | 132 | 5.1 | 1300 | 2 032591 | 032591 escherichia |
| 35 | 132 | 5.1 | 1300 | 2 032555 | 032555 escherichia |
| 36 | 131.5 | 5.0 | 319 | 5 077449 | 077449 heterodera |
| 37 | 130.5 | 5.0 | 319 | 5 061595 | 061595 heterodera |
| 38 | 130 | 5.0 | 392 | 5 044078 | 044078 globodera r |
| 39 | 129 | 4.9 | 391 | 5 077094 | 077094 globodera r |
| 40 | 129 | 4.9 | 519 | 2 047916 | 047916 fibrobacter |
| 41 | 129 | 4.9 | 1070 | 2 053365 | 053365 synechocyst |
| 42 | 129 | 4.9 | 1107 | 2 P76088 | P76088 escherichia |
| 43 | 129 | 4.9 | 1167 | 2 P76859 | P76859 escherichia |
| 44 | 129 | 4.9 | 1536 | 2 048031 | 048031 haemophilus |
| 45 | 127.5 | 4.9 | 739 | 2 09X687 | 09X687 salmonella |

ALIGNMENTS

| RESULT | 1 | PREDIMINARY: | PRT: | 516 AA. |
|--------|----------------------------------------------------------------------|--------------|------|---------|
| 066185 | AC | 066185; | | |
| ID | 066185 | PREDIMINARY: | PRT: | 516 AA. |
| DT | 01-AUG-1998 (TREMBLrel. 07, Created) | | | |
| DT | 01-AUG-1998 (TREMBLrel. 07, Last sequence update) | | | |
| DT | 01-MAR-2000 (TREMBLrel. 13, Last annotation update) | | | |
| DE | MANNANASE. | | | |
| OS | Bacillus circulans. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| CC | Bacillus/staphylococcus group; Bacillus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE; 98233274. | | | |
| RA | Yoshida S., Sako Y., Uchida A.; | | | |
| RT | "Cloning, sequence analysis, and expression in Escherichia coli of a | | | |
| RT | gene coding for an enzyme from Bacillus circulans K-1 that degrades | | | |
| RT | guar gum." | | | |
| RL | Biosci. Biotechnol. Biochem. 62:514-520(1998). | | | |
| DR | EMBL; AB007123; BAA23878.1; -. | | | |
| DR | INTERPRO; IPR001547; -. | | | |
| DR | PFAM; PF00150; cellulase; 1. | | | |
| SO | SEQUENCE 516 AA; 55245 MW; DBE5A48842AE8062 CRC64; | | | |

| | | | | |
|-----------------------|------------------|-----------------------------------------------------------|------------|-------------|
| Query Match | 62.28; | Score 1622; | DB 2; | Length 516; |
| Best local Similarity | 59.68; | Pred. No. 1.2e-107; | | |
| Matches 292; | Conservative 86; | Mismatches 84; | Indels 28; | Gaps 2; |
| QY | 27 | TSTNANGSGFVSGTTLTDANGNPFVNGINHGAMWDOATTIEGTANGANTVRVL | 86 | |
| DB | 29 | TNRKHAASGFVSGTKLDATGQPVKMGVNHATWYDQSLTAIPAKTGANTRIVL | 88 | |
| QY | 87 | SDGQWTKDQIHVTYNNLSLAEDNHLVAVPVVDATGYSIASINRAVDYIEKRSALIG | 146 | |
| DB | 89 | ANGHKWTDVNTVNNITLLOQNKLTAVLEVHATGSDSLSDNANNVYIGKSAIG | 148 | |
| QY | 147 | KEDVYIINAEWFGSGWEGDAMADGYKQALPRLNAGINHTLWADAGWGCFPOSIDHYG | 206 | |
| DB | 149 | KEDRYIINAEWYGTWGVAMANGYKQALPRLNAGITHTLWADAGWGCFPOSVKNYG | 208 | |
| QY | 207 | REVNADQQRNTWFSIHNYEYAGNAGSVNRINDIRVNLQDLALVIGEGHRTNGDVDEA | 266 | |


```

DB 209 TEVLNADPLAKNTVFSTHMEYAGNASTYKSIDVLANKNLLTIGEGCGHTNGDVDER 268
QY 267 TMTSTSEORGVGMWAMSKNGKPEWEYIDLSNDWAGNNLTANGNTVNGPYGLARETSRLS 326
DB 269 TMTSTSEORGVGMWAMSKNGKPEWEYIDLSNDWAGNNLTANGNTVNGPYGLARETSRLS 328
QY 327 TMTSTSEORGVGMWAMSKNGKPEWEYIDLSNDWAGNNLTANGNTVNGPYGLARETSRLS 360
DB 329 GTF--GGTTPYSSPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPTSTPT 386
QY 361 GPMATWEMSSKSGSLKADIDLSNSOHIATYIQTNTSLQONRIQATYKHAMGSGVNGM 420
DB 367 GPMATWEMKATGAQTLKADVLSQNSHSLTITSNOMNSKSLKATYKHAMGSGVNGM 446
QY 421 TARTLVKGGHTWYSGSVVPFNGSSGCTTSLDLSNVOMLSOVRIGYVQFOSGSSGCT 480
DB 447 YAKLVKGGHTWYSGSVVPFNGSSGCTTSLDLSNVOMLSOVRIGYVQFOSGSSGCT 506
QY 481 STYIDNVIVE 490
DB 507 AIYDVSLSQ 516

```

```

RESULT 2
ID 069347 PRELIMINARY; PRT: 395 AA.
AC 069347;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
GN MANA.
OS Vibrio sp.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-138;
RA Tamara Y., Ataki T., Morishita T., Kimura T., Sakka K., Ohmura K.;
RL J. Ferment. Bionng. 83:201-205(1997).
DR EMBL: D86329; BAA25188.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFM: PFO0150; cellulase; 1.
DR PFM: PFO2013; CBD_5; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 395 AA; 43098 MW; A3E30B3A978C5DA1 CRC64;

```

Query Match 31.28; Score 814.5; DB 2; Length 395;
 Best Local Similarity 46.58; Pred. No. 2.3e-50;
 Matches 168; Conservative 62; Mismatches 104; Indels 27; Gaps 9;

```

QY 6 KIKSITLSTLLASSILFVSGSTANANSGFYVSGTITDANGNPFYKNGINHGHWYKD 65
DB 5 KAIKSL-----LFTWASC--AAAGFYVSGVLYTEANGSAFKRGINHAHTWTD 52
QY 66 QATTAIEGANTGANTVRYVLSGCGWTKDDITVYRNLSIAEDNHLVAVEVDATGY- 124
DB 53 KLSVALSGIATGANTVRYVLSGCGWTKDDITVYRNLSIAEDNHLVAVEVDATGY- 112
QY 125 --DSIASLNRAVDYWIEMRSALIGKEDVTIINANWFG--SWEGDANAGYKQAIPLRN 181
DB 113 EESSAASIDSADYWIEMRSALIGKEDVTIINANWFG--SWEGDANAGYKQAIPLRN 172
QY 182 AGLNHTLVNDAAGWQ--FPOSIDYGREVNADPQRTMSEIHME--YAGNASQVRENT 239
DB 173 AGLNHTLVNDAAGWQ--FPOSIDYGREVNADPQRTMSEIHME--YAGNASQVRENT 230
QY 240 DRYLNODLALVYIGFGRHTNGVDVDEATIMSYSEORGVGMWAMSKNGKPEWEYIDLSND 299
DB 231 SSFTNNGLVLYIGFGRHTNGVDVDEATIMSYSEORGVGMWAMSKNGKPEWEYIDLSND 290

```

```

QY 300 WAGNNLTANGNTVNGPYGLARETSRLSTVFTGG----SDGCTSP---TLYDEFGSMOG 352
DB 291 WDNNSYSTWGMVLLINGONGIKSTSLTAVFTCGNDCCDDSSGGEYPCSSAVDPDGGWG 350
QY 353 W 353
DB 351 W 351

```

```

RESULT 3
ID 086599 PRELIMINARY; PRT: 384 AA.
AC 086599;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE SECRETED BETA-MANNOSIDASE.
GN MANA2.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Batteil B.G., Rajandream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL031514; CA20610.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFM: PFO0150; cellulase; 1.
DR PFM: PFO2013; CBD_5; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 384 AA; 39847 MW; B8AD158E3FD3BD11 CRC64;

```

Query Match 29.28; Score 761.5; DB 2; Length 384;
 Best Local Similarity 44.28; Pred. No. 1.3e-46;
 Matches 161; Conservative 59; Mismatches 121; Indels 23; Gaps 8;

```

QY 9 FSITLSTLLASSILFVSGSTANANSGFYVSGTITDANGNPFYKNGINHGHWYKD 65
DB 16 FAAYLGLFLA-----LAGSAGRAEAAAGSIHNSNGVLENGSVYVRGNNATWYTPD 70
QY 66 QATTAIEGANTGANTVRYVLSGCGWTKDDITVYRNLSIAEDNHLVAVEVDATGY- 124
DB 71 R-TGSINADIAKANTVRYVLSGCGWTKDDITVYRNLSIAEDNHLVAVEVDATGY- 129
QY 125 --DSIASLNRAVDYWIEMRSALIGKEDVTIINANWFG--SWEGDANAGYKQAIPLRN 182
DB 130 EDGAASISLDAADYWIEMRSALIGKEDVTIINANWFG--SWEGDANAGYKQAIPLRN 189
QY 183 GLNHTLVNDAAGWQ--FPOSIDYGREVNADPQRTMSEIHME--YAGNASQVRENT 241
DB 190 GLNHTLVNDAAGWQ--FPOSIDYGREVNADPQRTMSEIHME--YAGNASQVRENT 248
QY 242 VLNODLALVYIGFGRHTNGVDVDEATIMSYSEORGVGMWAMSKNGKPEWEYIDLSND 301
DB 249 FVSGSLPIYVIGFGRHTNGVDVDEATIMSYSEORGVGMWAMSKNGKPEWEYIDLSND 308

```


09RJ15 PRELIMINARY; PRT; 327 AA.
 AC 09RJ15: TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
 DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
 GN MANA.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parthill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleiser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL121746; CAB57406.1; -.
 DR INTERPRO: IPR001547; -.
 DR PFM: PFO0150; cellulase; 1.
 FT NON_TER 327
 SQ SEQUENCE 327 AA; 34772 MW; A21B068C2C0975D CRC64;

Query Match 20.8%; Score 541; DB 2; Length 327;
 Best Local Similarity 41.5%; Fred. No. 5e-31;
 Matches 117; Conservative 54; Mismatches 101; Indels 10; Gaps 8;

QY 25 SGGSTANAN-SGFFVSGTLLYDANGNPFVARGINCHGAWTKDOATTAIEGANTGATVA 83
 DB 47 AGPSSAGAGAGAGHIGDGRLEBNGDNFVARGYNHATWYRPGF-TQSLADYKALGANSVA 105
 QY 84 IVLSDGQWTKDIDHYVRLNLSLAEDNHLVAVEVDATGY--DSTA-SINRAVDYIEM 140
 DB 106 VVLSDDHRRGSENGPADVAAYIEQCKANRLICYLEVHDTGTGAEDAAAGTIDHADYWIGL 165
 QY 141 RSLALIGEDVIYININEMGSGWEGDAMADGYQALPRLNAGINHTLMDADAGWGQFPQ 200
 DB 166 KDVLAQGEDVIYINIGEPNGNDPAGWTEPTVAAYKLLAAGLQHTIMVADAPWNGDQW 225
 QY 201 SI-HDYGREVENADPORHTESIMHYEAGNNSQVNTNIDRYLMDLALVIGEF-GHRH 258
 DB 226 GVMRANRSYVDADPTGNLFLFSIHMS-VFDTAQETIDYINAFVADALPILIGFEGPAD 284
 QY 259 TNGDVDEATMSTSEORGWGLAWSKNGGPEWETDLSMDW 300
 DB 285 QYSDPEDTMMATAEOLRLGYLAWMSGND--FVLDLADF 324
 RESULT 7
 09ZAI7 PRELIMINARY; PRT; 1097 AA.
 AC 09ZAI7: TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, last annotation update)
 DE S-LAYER ASSOCIATED MULTIDOMAIN ENDOGLUCANASE.
 GN CELA.
 OS anaerobic thermophile KM-THCT.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
 RN (1)

RP SEQUENCE FROM N.A.
 RC STRAIN-KM-THCT;
 RX MEDLINE; 99175466.
 RA Cann I.K., Kocherginskaya S., King M.R., White B.A., Mackie R.I.;
 RT "Molecular cloning, sequencing, and expression of a novel multidomain
 RT mannase gene from the anaerobic bacterium polysaccharolyticum";
 RL J. Bacteriol. 181:1643-1651(1999).
 DR EMBL; 082255; AAD09354.1; -.
 DR INTERPRO: IPR001119; -.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR002088; -.
 DR PFM: PFO0150; cellulase; 1.
 DR PFM: PFO0395; SLH; 3.
 DR PFM: PFO2018; CBD; 6; 2.
 DR PROSITE: PS00904; PETA; UNKNOWN_1.
 DR PROSITE: PS01072; SLH_DOMAIN; 2.
 SQ SEQUENCE 1097 AA; 119762 MW; C41CEB92C0494B9 CRC64;

Query Match 9.7%; Score 253.5; DB 2; Length 1097;
 Best Local Similarity 22.3%; Fred. No. 7.8e-10;
 Matches 104; Conservative 76; Mismatches 181; Indels 105; Gaps 20;

QY 8 IFSITSLIASILFVSGTSTANANSG--FYVSGTLLYDANGNPFVARGIN-HGH-AW 62
 DB 10 VVAVVLTVALNGVYIGDKQAAAGTSCDGRHYVGNKIYDPDGNDFYIKVNGIGYRSM 69
 QY 63 YDQATTAIEGANTGA-NVARIYLSG-----GQWTKDHYVRLNLSLAEDNHLVAV 115
 DB 70 ERSVYLDVHLADYWKFTYVRLNCFIGNMWSGEGTGANNDIAI--IKAFYAKVVE 126
 QY 116 PEVHATGYSIA-----SLNRAVDYIEMSGALIGKEDVIYINANWFGSWE-- 164
 DB 127 IDLHDTGTGPPLSNPPAPGQPSLDQAI-AWKEIAAYIKDNPIYFWNTMNEPSSAPL 185
 QY 165 GDAWADGYKQALPRLNAGINHTLMDADAGWGQFQSIHD-----YGREVF 210
 DB 186 DQWYANDEIKAIKRSAGDNIIVYD--GWSYANEGIEQNTPTDEKRSAYLTATGQDL 243
 QY 211 NADPQNTWESIMHYEAGNNSQVNTNIDRYLMDLALVIGEFHRTNGVDATIMS 270
 DB 244 NADSAKNTFFAHYNN-EDDIQKVEDYIDRANAGLYVMEYGRKDY--DAAKEGVKS 300
 QY 271 YSE---QRGVGLAWSKNGGPEWETDLSNDWAGNNLTMAGNTLVNGPYGLRSTSLST 327
 DB 301 GLQAVNNKAGRIYNNWDS---YDYLDTSG----- 328
 QY 328 VFTGGS-----DGTSPFTLYDEGSMQGTGSLSGPVAVTEWSSKGSLSKADIQ 381
 DB 329 --TGRSGWEIKKTGSKRTNL-----SWVGDKIMDNDHGT--PTFDONPKVDLA 376
 QY 382 LSSNSOHLVHVIQNTSLQONSRIQATVKRANGSGVNGHTALYK 427
 DB 377 LER-----LIANNNGFKAGDRVOFTFLRNSGDLPIGDSKRYVK 416

RESULT 8
 P94622 PRELIMINARY; PRT; 557 AA.
 AC P94622: TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, last sequence update)
 DT 01-MAY-1997 (TREMBlrel. 03, last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, last annotation update)
 DE ENDO-1,4-BETA GLUCANASE ENGF (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
 DE (ENDO-1,4-BETA-GLUCANASE) (CANBOXYMETHYL CELLULOSE).
 GN ENGF.
 OS Clostridium cellulovorans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97136706.
 RA Shewetta S.A., Ichl-Ishl A., Park J.S., Liu C., Malburg L.M.,

RA Doi R.H.:
 RT "Characterization of enf, a gene for a non-cellulosomal Clostridium
 RL cellulovorans endoglucanase";
 CC Gene 162:163-167(1996).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: U37056; AAB40891.1; .
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; .
 DR PFMW: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Hydrolase; Glycosidase.
 SO SEQUENCE 557 AA; 60131 MW; D186EC88E504EED CRC64;

Query Match 9.2%; Score 239.5; DB 2; Length 557;
 Best Local Similarity 23.1%; Pred. No. 3e-09;
 Matches 139; Conservative 74; Mismatches 221; Indels 169; Gaps 29;

QY 2 NNGFKRIFSI-----FLSLASSILFVSGTS-----TANM-----SGFTV 38
 DB 3 NNVKRRILSIYAGAMLMALVPINVAETYSNLTGNVKKPSVGRKQLQINKGI-- 60
 QY 39 SCTLYDANGNPFVNRGIN-HGHAWTKDQATAIIEGIANTGA-----NTVRIY- 86
 DB 61 --KTLCDKDNFQOLRGKSTHGLQNF-----PQVNNNAFALSDNWSNVITLAMY 110
 QY 87 -SDGQWTKDDI-HVYRNLSIAEDNHLVAVPEVDATGYDSIASINRAVDYWIERSAL 144
 DB 111 VAEQVATNPSVKQVINGINVAIANDMVYIVDMHMGNDNAGVYSQAQSFNDISLT 170
 QY 145 ICKEPTVITINAEWFGSEEG--DA--WADGYKQALP--RLRNAGLHETLMDAAGCG 196
 DB 171 YPNNNIITYELCNEPENGNGVGTNDATGWAQYSATPIVQLLRKNGENLITVGNPEWS 230
 QY 197 QEPQSIHDGREFVNAADPQRTMFSIHMYEYAGN-ASQVETNIDR-----VLNQDA 248
 DB 231 QRPDLAAD-----NPINDSTMTSYHF--YSGTNPISIVTDNRDNAMSNVATYALNHGA 282
 QY 249 LVIGEGHRTNGDVEDATIMSYSEQGVGLAWSKNGSPMEYLDLSND----- 299
 DB 283 VFATEMGTSLATG-----TTGPYL-AKADAWLDF-LNGNNISMCMNPSINKDEKAALNS 335
 QY 300 -----WAGNNLTAMG-----NTIVGPGGLRRTSKLSTVFTGG 332
 DB 336 LTSIDPGSDKPLADNELTTSQGYVARIKGAYYAPVPVTQPAKDFSSGFMDFNQ 395
 QY 333 GSDG-----GTSPTLYDFEGSGMGWGTGSLSGPNAVTEWSKSGSHLK-----ADIOS 383
 DB 396 TTQGGVNPDSFIYALINVENANNALKISML-----NSKGSNDLSEGNFMANVRIS 445
 QY 384 SNS-OHYLVAVIONTSLOQSRIOATVYKRAHMGVSGNGMTARLYVYKGGHY-----TW 434
 DB 446 ADIMGOSINIYDGTKLMDVIAPTPV-----NYSIALPOSTHGWGPTAIRVW 496
 QY 435 YSGSFVPIINGSGCTLSLSDLSNVQLSQVREIGVQFQASDS-----SGQTSITI 484
 DB 497 TNNFVAQTDGYKATLTISTNDSPNFTIA-----TDAADSVYTMILFVSSNDISLT 550
 QY 485 DNV 487
 DB 551 DNI 553

RESULT 9
 ID 059232 PRELIMINARY; PRT; 389 AA.
 AC 059232
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ENDO-BETA-1,4-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULASE)
 DE (ENDOGUCANASE) (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE).

OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 RL Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-186-1;
 RA Sanchez Torres J.;
 RL Thesis (1994), Universidad de Salamanca, Spain.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: 233876; CA83942.1; .
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; .
 DR PFMW: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR KW Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 25
 FT CHAIN 26 389
 SO SEQUENCE 389 AA; 43688 MW; 9111397485609B5F CRC64;

Query Match 8.7%; Score 228; DB 2; Length 389;
 Best Local Similarity 23.6%; Pred. No. 1.2e-08;
 Matches 100; Conservative 65; Mismatches 154; Indels 104; Gaps 21;

QY 6 KRFISITSLASSILFVSGTSTANANS-----GFYVSGTLYDANGNPFVNRGI-NH 58
 DB 2 KRITIFVLLMTVALFSIGNTTAADNSVVEHGQLSISNGELVNERGEQVQLGMSH 61
 QY 59 GHAWTKDQATAIIEGANTGATVRIYLSDGQWTKDD-----IHT----- 99
 DB 62 GLQWVG-----QFVNESKMLRDLMDGIVFRAAMTSSGGYIDPS 103
 QY 100 ---YRNLSIAEDNHLVAVPEVDATGYDSIASINRAVDYWIERSALIGKEDYIINI 155
 DB 104 VAEKTKAEVAEALDDIYIIDMHLSDNDPRYIKERKDFDEM-SELYGYPVYIYI 162
 QY 156 ANEWFGSEGDAMADGYK---QALPRLNAGLHETLMDAAGWQFPOSIDHYREVFN 211
 DB 163 ANEPGSS--DYWDNRIRKYAEVIVPIRNNDPNNIIV---GTQWSDVH-HAADNL 216
 QY 212 ADPQRTMFSIHMYEYAGNAGVQVETNIDRVNLQDLAVIGEGHRTNGD-----YDEA 266
 DB 217 ADP--NVMTAFHF--YAGTREGONLDQVDYALDQGAALFVSWGSAATGDCGVLEDEQ 272
 QY 267 TIMSYSEQGVGLAWS-----WKGNGPMEYLDLSNDVAGNNLTAMGNTIYNG 315
 DB 273 EVIDMDERNLSNMAWSLTHKDESSAALMPGANP-----TGWTAEALSFGTFVRE 324
 QY 316 PYGLRSTSLSTVFTGGSDGT---SPTLYDFE-----GSM---OGWTGSSLSG--- 360
 DB 325 --KIRSEASIPSPDPTPPEDPGEYPAMPPTQYLTNEIYVHNGQLWQAKWTQNOGPDY 382
 QY 361 GPW 363
 DB 383 GPW 385

RESULT 10
 ID 059665 PRELIMINARY; PRT; 570 AA.
 AC 059665
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGUCANASE)
 DE (CARBOXYMETHYL CELLULASE).
 GN CELE.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 CC Pseudomonas.
 RN [1]
 SO SEQUENCE FROM N.A.

RX MEDLINE: 95366948.
 RA Hall J., Black G.W., Ferreira L.M.A., Millward-Sadler S.J.,
 RA Ali B.R.S., Hazlewood G.P., Gilbert H.J.,
 RT "The non-catalytic cellulose-binding domain of a novel cellulase from
 RT Pseudomonas fluorescens subsp. cellulosa is important for the
 RT efficient hydrolysis of Avicel."
 RL Biochem. J. 309:749-756(1995).
 CC -1- CATALYTIC ACTIVITY: ENOCHDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 DR EMBL: X86798; CAA60493.1; -
 DR HSSP: P07103; 1EG2.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001979; -
 DR INTERPRO: IPR002883; -
 DR PFAM: PF00150; cellulase: 1.
 DR PFAM: PF00553; CBD_2; 1.
 DR PFAM: PF02013; CBD_5; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR Hydrolase: Glycosidase.
 KW SEQUENCE 570 AA; 59925 MW; D00B26F13D57FBE7 CRC64;

Query Match 7.8%; Score 203.5; DB 2; Length 570;
 Best Local Similarity 23.3%; Pred. No. 1.1e-06;
 Matches 119; Conservative 75; Mismatches 197; Indels 119; Gaps 26;

QY 14 SLILASSILEFSGSTANANGSGFYGGTLLYDANGNPFVARGIN-----HGAMYED 65
 DB 22 AIIILSFLGLGVSNAQADVAPLSVGNKIL-ANGOPASPGSLFWSMTEMGEXYNA 80
 QY 66 QATTAIEGANTGANTVRIY---SPGGOWT-----KDDITVYNNLSIAEDNHLVAVPY 118
 DB 81 QVVSFWLK--SDMNAKLYRAMAGVEDEGGYLPDANKDRVYQVVD-AALADMTYIIDMS 137
 QY 119 HDATGDSIASLRAYDWTIEMRSALIGKEDTYINIANEMFG-SWEGDAMADYKCA-I 176
 DB 138 HNHQJQS-----QALFEQEM-ARKYGANNHYIETINEPLVYMSNT--IKPYQAVI 189
 QY 177 PRLNAGLINTLWADAGMGOFPQSIHDYGEVFNADP---QRTNMFSLIMTYAGNMS 233
 DB 190 AAIIRAIDPMLIIVGTPTWSQ-----DVDVANDPITGYONIAATLHF--YAGHGQ 239
 QY 234 QVETNIDRYANQDLALYIGFEGHRTNGDVEDNTIMSYSPQRGVGLAMWKGMPWEX 293
 DB 240 YLRKQACTALNRGIALFTVTEGVSANNGD-----GAVANSE-----GW 315
 QY 294 LIDLSNDVAGNNLITAMGNTIYNGPYGLRETSRLSTVFTGGSGDGTSPITLYDEGSMQW 353
 DB 276 ---TNANVSFKYT---KHISNANALNDKYEGASALVPGASANG-----GW 315
 QY 354 TGSSL-SGGPWA---VTEW---SSKGSLSLKADQLSSNSQHTLHYQNTSLQONSRIQ 405
 DB 316 VNSQLTASGALAKSLIISQWPSYNTSSSSSAVSQTOVSSSQ--APVSSSSSTASSVVS 373
 QY 406 ATY--KHAMGSGVNGMTARLYVKTGHGYT-----YSGSFVPIGSSGTTLSL 452
 DB 374 SAVSGGQCNM-----YGLIYPLCSTTTNGMGWENNASCIARATCGGAPAPMGIVGSSIS- 428
 QY 453 DLSNVQNLQVREIGVQFQASDSGQTSI 482
 DB 429 -----QASSSVRSSSSSLVSSRSSSSSV 453

RESULT 11
 Q45532 PRELIMINARY; PRT; 499 AA.
 AC Q45532;
 DT 01-JAN-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE CELULASE.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87190397.
 RA Nakamura A., Uozumi T., Teruhiko B.,
 RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis."
 RL Eur. J. Biochem. 164:317-320(1987).
 DR EMBL: M28332; AA422307.1; -
 DR HSSP: O85465; 2A3H.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00150; cellulase: 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -; 1.
 SQ SEQUENCE 499 AA; 55075 MW; D42AAB9E63B94D23 CRC64;

Query Match 7.4%; Score 193; DB 2; Length 499;
 Best Local Similarity 20.5%; Pred. No. 5.2e-06;
 Matches 105; Conservative 81; Mismatches 165; Indels 160; Gaps 26;

QY 6 KRIFSTILSLASSILEF-----VSGSTANANGS-FYVSGTLLYDANGNPFVARG 55
 DB 2 KRISIFITCLITVLTWGLQASPASAGTKTPAAKNGQISLGTOLVNDGKRAVOLKG 61
 QY 56 I-NHGAMYKQOATTAIEGANTGANTVRIYLSPPGGOWTKD-----IHT----- 99
 DB 62 ISSHLOWYGD-----FYNNLSIAEDNHLVAVPEVDATGYDSIASLRAYDWTIEMRSALIGKEDTY 103
 QY 100 -----VRNLSIAEDNHLVAVPEVDATGYDSIASLRAYDWTIEMRSALIGKEDTY 151
 DB 104 DNPSTKNNKVEAKELGIYIIDLNDGNPQKEREKDFEEM--SLGLGNPNV 162
 QY 152 IINIANEMFG--SWEGD--ANADGYKQAIPLRRAGNHTLWADAGMGOFPQSIHDYGR 207
 DB 163 IYEIANENPGVNMWRDKRPAE--EVIYVIRKNDPNTITV--GTVWSQVND-- 213
 QY 208 EYFNADPO---RNTMFSLIMTYAGNMSQVETNIDRYANQDLALYIGFEGHRTNGD-- 262
 DB 214 ---AADOLKANVYALHF--YAGTHQSLRKNVYALSGLAPLFTVEWGTSDASGNG 268
 QY 263 ---VDEATIMSYSQRGVGLAMWKGMPW--EYIDLSN--DWAGNNLITAMGNTIYNGPYG 318
 DB 269 VFLDQSR-----EWLNTYDSKNISVYNNL----- 293
 QY 319 LRETSRLSTVFTGGSGDGTSPITLYDEGSMQWTSLSGGPAPVTEMSKGSLSKA 378
 DB 294 -----SDKOESSALK-----PGASKTGG-WPLTDLTASGTF-VRE 327
 QY 379 DI--QLSSNSQHTLHYQNTSLQONSRIQATVYKHAMGSGVNGMTARLYVKTGHGYT--- 433
 DB 328 NILGNNDSTKREPERPAQONPAENG-IEYQYKADGDGVNSQIHPOLHIKKNGNATYDL 386
 QY 434 -----WTSGSFVPIGSSGTTLSIDLSNV 457
 DB 387 KDVTARYWYNNAK-----NKGQNFDDYAOI 411

RESULT 12
 O18453
 ID O18453 PRELIMINARY; PRT; 476 AA.
 AC O18453;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE BETA-1,4-ENDOGLUCANASE-1 (CARBOXYMETHYL CELLULASE).
 DE (ENDO-1,4-BETA-GLUCANASE)
 GN HG-ENG-1.
 OS Heterodera glycines.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 98226740.
 RA Smart G., Stokkermans J.P.W.G., Yan Y., De Boer J.M., Baum T.J.,
 Wang X., Huseey R.S., Gommers F.J., Henriksat B., Davis E.L.,
 RA "Bakker J., Schots A., Bakker J.,
 RA "endogenous cellulases in animals: isolation of beta-1, 4-
 RT endoglucanase genes from two species of plant-parasitic cyst
 nematodes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4906-4911(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OP50;
 RA Yan Y., Smart G., Stokkermans J.P.W.G., Wang X., Huseey R.S.,
 RA Bakker J., Helder J., Schots A., Davis E.L.,
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC "Catalytic Activity: ENDOPHOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE."
 DR EMBL: AF006052; AAC48327.1;
 DR EMBL: AF052733; AAC15707.1;
 DR HSSP: P07103; 1RGZ
 DR INTERPRO: IPR001547;
 DR PFAM: PF00150; cellulase, 1.
 DR Signal: Hydrolase; glycosidase.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 476 BETA-1,4-ENDOGIUCANASE-1.
 SQ SEQUENCE 476 AA; 49815 MW; 2056F90CA0783FF7 CRC64;

Query Match 7.2%; Score 187.5; DB 5; Length 476;
 Best Local Similarity 23.7%; Pred. No. 1.2e-05;
 Matches 120; Conservative 61; Mismatches 211; Indels 115; Gaps 23;

QY 13 LSLILLASLIFVSG-TSTANANSGFYSGTLLDANGNPFYMRGIN-HGHAMTKD-CAT 68
 DB 13 LELLALCTALYSLLTAVAPFGOLSVGTMLVYGANGQPOLLISLFWHMYPOFWMAD 72
 QY 69 TAISGANTGANVYR--IVLSGGGWTGKDDIHYRNT-----ISLAEDNHLVAVPEVHDAT 122
 DB 73 TVKALKCKMNNANVIRGANGVDEGGTL--DANTAVNLMVAIVTAISNGITYIVDMHANN 130
 QY 123 GYDSIASINRAVDYWIEMRSALIGEDTVIINIANEMFG-SWEGDAMADGYKQALPRLN 181
 DB 131 SHP-----DEAVKFFTRTAQ-YGSYPHILIEDPEPESVEMT-DVLYPIHKVIAALRA 183
 QY 182 AGLNHTLMDAAGGQ-----FPOSIDHYGREVNADFORNTWFSIHMYETAGNASQVR 236
 DB 184 IDKNVILIGTPWSDVDVASQNPINDY-----QNLMTLHFY-----ASSHF 227
 QY 237 TN-----IDRLNDLALVIGFEG--HRTNGDDVDATINSYSEKQYGV-LAMSWKNG 288
 DB 228 TNDLGAKIKTVANNGLPVFVTEYGECAASGNLMTDSMSW-----WTLLSLKLSY 280
 QY 289 PEMEYLDLS-----NDMAGNNLTAMGNTIYNGPYGLBETSRSTVFTGGSD 335
 DB 281 ANMAISDEKSCALSPTTLAANGVSSRWTSNGMVAS--YTKKSGVSCSGSSGSS 338
 QY 336 GGTSPITLLYDEGSMQGTGSSLSGPPWAVTEWSSKSHSKADQLQSSNOHYLAVION 395
 DB 339 SGSSSGSSGSSGSSGSSGSSGSSG-----SSGSSGSSGSSGSSGSSGSSGSSGSS 390
 QY 396 TSLQONSHIQATYVKAHANGSYGNGTAAIYKGTGHGTYWYSGFYPINGSGTTLISLDS 455
 DB 391 T--WNGGGVNFPIKNGSVP-----LCGVVFVSILPSGTL----- 425
 QY 456 NVONLSQVREIGVOFQASDSSGOTS 482
 DB 426 -----GGSMNESAGSGOTSL 441

RESULT 13
 ID 066064 PRELIMINARY; PRT: 481 AA.
 AC 066064;

DT 01-AUG-1998 (TREMBREL. 07, Created)
 DT 01-AUG-1998 (TREMBREL. 07, Last sequence update)
 DT 01-JUN-2000 (TREMBREL. 14, Last annotation update)
 DE ENDOGLUCANASE.
 OC Actinomyces sp. 40.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-40;
 RA CHO K.K.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U94825; AAC06196.1;
 DR HSSP: O85465; 2A3H
 DR INTERPRO: IPR001547;
 DR INTERPRO: IPR001919;
 DR PFAM: PF00150; cellulase, 1.
 DR PFAM: PF00553; CBD_2; 1.
 DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; UNKNOWN_1.
 SQ SEQUENCE 481 AA; 50750 MW; 50A6A29BEE44057A CRC64;

Query Match 7.2%; Score 187.5; DB 2; Length 481;
 Best Local Similarity 20.6%; Pred. No. 1.2e-05;
 Matches 100; Conservative 76; Mismatches 212; Indels 97; Gaps 21;

QY 38 VSGTLLDANGNPFYMRGIN-HGHAMTKDQAT-TAISEGANT-GANTYRIVL-----S 87
 DB 45 ISAPFTIIDEHGMPPOLNRASHTHGLQWGYNKAQFSLRDEMGLNRLVAVYREGYL 104
 QY 88 DGGGWTGDDIHYRNTLSLAEDNHLVAVPEV-----DATGYDSIASINRAVDYWI 139
 DB 105 QGSAQMD--KTIOGVQAATDGLMIVYIIDHVLNPNNGDATQAES-----PFK 152
 QY 140 MRSALIGEDTVIINIANEMFGSWEGDAMADGYKQAL-----PRLNAGNHTLMD 191
 DB 153 SYAKKYSYGVIVCEVCEPCTGT-----PWYDGSNDIYSCTHMAKAIROAGDAIILCG 208
 QY 192 AAGNGQFPOSIDHYGREVNADFORNTWFSIHMYETAGNASQVRNTDRYLNDDLALVI 251
 DB 209 TINTWS--QDIDAVAGKPLSADGFDNIMYLVH--YATHDDIDRAKLQTLALNGTPVFV 263
 QY 252 GEGF--HRTNGDDVDENI--MSYSEKQYGVGLAMSWKNGPMEY-----LDSNDWAG 302
 DB 264 SEFGICDASGGGIDQDSANAMMTLLAHNNISYAMALSNKAEITRAFPSTVATSKWTG 323
 QY 303 NNLTAMGNTIYNGPYGLBETSRSTVFTGGSDGTPSTLLYDEGSMQGTG--SSLSG 360
 DB 324 DDLTPSAIMLVNTSRKLADPADHAASGTSAGSSRAAGSGTGAGNGKSSASASSPSAAG 383
 QY 361 GPWAV--TWSKSHSLKADQLQSSNOHYLAVIONTSIQONSRIQATVKAHANGSV 416
 DB 384 GLTFATAIRNOMNGATVAIL--TVSNASGSKH-----EGAMQVTFPDADIADI-WG-- 432
 QY 417 GNGMTAALYVGTGTYTSSGFYPINGSGTTLISLDSNVONLSQVREIGVOQASADS 476
 DB 433 -----GTIVSHGTHYVAPMDMNTAIEAGASAEIIGN--ASS 468
 QY 477 SGOTS 481
 DB 469 TGOAS 473

RESULT 14
 ID 083012 PRELIMINARY; PRT: 501 AA.
 AC 083012;
 DT 01-NOV-1998 (TREMBREL. 08, Created)
 DT 01-NOV-1998 (TREMBREL. 08, Last sequence update)
 DT 01-JUN-2000 (TREMBREL. 14, Last annotation update)
 DE CELLULOSE.
 OS Bacillus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:49:31 ; Search time 214.78 Seconds
(without alignments)
47.761 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 MANSGRYVSGTLTYDANGNP.....TIVNGPYGURETSRSTVFT 300

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 1630 | 100.0 | 490 | Y54122 | Amino acid sequenc |
| 2 | 1624 | 99.6 | 476 | Y54123 | A mannanase-linker |
| 3 | 1283 | 78.7 | 468 | Y44496 | Bacillus agaradher |
| 4 | 1283 | 78.7 | 468 | Y54125 | Amino acid sequenc |
| 5 | 1283 | 78.7 | 493 | Y44495 | Bacillus agaradher |
| 6 | 1283 | 78.7 | 493 | Y54124 | Amino acid sequenc |
| 7 | 1136 | 69.7 | 331 | Y54127 | Amino acid sequenc |
| 8 | 1075 | 66.0 | 369 | Y54129 | Amino acid sequenc |
| 9 | 937.5 | 57.5 | 320 | Y54132 | Amino acid sequenc |
| 10 | 522 | 32.0 | 188 | Y54133 | Amino acid sequenc |
| 11 | 369 | 22.6 | 305 | Y54130 | Amino acid sequenc |
| 12 | 316 | 19.5 | 132 | Y54131 | Amino acid sequenc |

| | | | | | |
|----|-------|------|------|----|--------|
| 13 | 202 | 12.4 | 410 | 18 | W12378 |
| 14 | 202 | 12.4 | 411 | 18 | W12379 |
| 15 | 202 | 12.4 | 411 | 18 | W12381 |
| 16 | 202 | 12.4 | 412 | 18 | W12380 |
| 17 | 201 | 12.3 | 400 | 18 | W23601 |
| 18 | 201 | 12.3 | 400 | 18 | W22521 |
| 19 | 201 | 12.3 | 400 | 19 | W57431 |
| 20 | 201 | 12.3 | 462 | 19 | W57433 |
| 21 | 200 | 12.3 | 409 | 9 | P81843 |
| 22 | 191 | 11.7 | 467 | 17 | W05731 |
| 23 | 191 | 11.7 | 467 | 17 | W00382 |
| 24 | 174.5 | 10.7 | 499 | 14 | R42122 |
| 25 | 171 | 10.5 | 352 | 20 | P08472 |
| 26 | 156 | 9.6 | 1010 | 19 | W34989 |
| 27 | 152 | 9.3 | 800 | 8 | P70420 |
| 28 | 152 | 9.3 | 822 | 13 | R26021 |
| 29 | 149.5 | 9.2 | 357 | 16 | R77394 |
| 30 | 149.5 | 9.2 | 941 | 11 | R07478 |
| 31 | 149.5 | 9.2 | 941 | 16 | R77395 |
| 32 | 142 | 8.7 | 358 | 19 | W39362 |
| 33 | 142 | 8.7 | 521 | 17 | R89827 |
| 34 | 142 | 8.7 | 562 | 21 | V69508 |
| 35 | 135 | 8.3 | 551 | 14 | W18790 |
| 36 | 127 | 7.8 | 472 | 19 | W37243 |
| 37 | 126 | 7.7 | 302 | 19 | W43909 |
| 38 | 124 | 7.6 | 484 | 19 | W43910 |
| 39 | 118.5 | 7.3 | 956 | 19 | W49874 |
| 40 | 118 | 7.2 | 476 | 19 | W37241 |
| 41 | 112.5 | 6.9 | 360 | 19 | W49870 |
| 42 | 112.5 | 6.9 | 386 | 19 | W37242 |
| 43 | 111.5 | 6.8 | 360 | 18 | W34565 |
| 44 | 110.5 | 6.8 | 532 | 12 | R13329 |
| 45 | 110 | 6.7 | 531 | 16 | W01503 |

ALIGNMENTS

| | | |
|----------|--------------------------------------------------------------------------|-------------------------------------|
| RESULT 1 | Y54122 | standard; Protein; 490 AA. |
| XX | ID | Y54122; |
| XX | AC | Y54122; |
| DT | 27-MAR-2000 | (first entry) |
| XX | XX | |
| DE | Amino acid sequence of a Bacillus mannanase enzyme. | |
| XX | XX | |
| KW | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; | |
| KW | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; | |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan; | |
| KW | galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; | |
| KW | printing paste; plant material degradation; recycled waste paper; | |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; | |
| KW | mannan-containing food; coffee extract; cleaning composition; | |
| KW | machine washing; hard-surface cleaner; dishwashing; oral; dental; | |
| KW | contact lens; body-care composition; fabric softener; oil well drilling; | |
| KW | subterranean formation fracture. | |
| XX | OS | Bacillus sp. |
| XX | XX | |
| FH | Key | Location/Qualifiers |
| FT | Peptide | 1..30 |
| FT | FT | /note- "signal peptide" |
| FT | Domain | 31..330 |
| FT | FT | /note- "catalytic domain" |
| FT | Misc-difference | 116 |
| FT | FT | /note- "encoded by CYP" |
| FT | Domain | 331..342 |
| FT | FT | /note- "linker" |
| FT | Domain | 343..490 |
| FT | FT | /note- "domain of unknown function" |
| XX | XX | |

P300-Ce1B fusion c
P300-Ce1B fusion c
P300-Ce1B fusion c
P300-Ce1B fusion c
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
Sequence of alkali
Cellulase. Bacill
Bacillus cellulase
NK-1 cellulase. B
Actinomycete sp. 3
Terebinthaceae end
Sequence encoded b
Alkaline cellulase
Fragment of alkali
Cellulase. Bacill
Full length Bacill
A. cellulolyticus
A. cellulolyticus
Acidothermus cell
Corrected Bacillus
Globodera rostoch
Globodera glycine
Heterodera glycine
Bankia goudi glyc
Heterodera glycine
Thermotoga OC1/4V
Thermotoga OC1/4V
Endoglucanase enco
60 kD endoglucanas

PN WO9964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kaupplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI; 2000-105891/09.
 DR N-PSDB; 245335.
 XX
 PT New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PS Claim 1; Page 208-210; 242pp; English.
 XX
 XX The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
 CC of *Bacillus* sp. 1633. The mannanase hydrolyses galactomannans.
 CC Specifically, mannanses hydrolyse 1,4-beta-D-mannosidic linkages in
 CC mannans, galactomannans, glucomannans, and galactoglucmannans. The
 CC mannanase protein, or preparations containing it, are used to improve
 CC properties of cellulosic or synthetic fibres, yarn or (non)woven
 CC fabrics (removal of mannan-based sizes or printing pastes). They are
 CC also used to degrade or modify plant materials (particularly recycled
 CC waste paper, paper making pulp, or material containing guar or locust
 CC bean gums (thickeners), or to reduce viscosity of mannan-containing
 CC foods (or feeds). The mannanses are also used to process coffee
 CC extracts (to inhibit gel formation); in cleaning compositions (for
 CC machine washing of fabrics, as hard-surface cleaners, for hand or
 CC machine dishwashing, also in oral, dental, contact lens or body-care
 CC compositions) where they remove mannan-containing soils and prevent
 CC binding of some soils to cellulosics; and in fabric softeners. They
 CC can also be used in oil well drilling to fracture subterranean
 CC formations.
 CC
 CC Sequence 490 AA:
 XX
 XX

Query Match 100.0%; Score 1630; DB 21; Length 490;
 Best Local Similarity 100.0%; Pred. No. 7.8e-141;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NANSFYVSGTLYDANGPFYRGINGHANYKDOATTAIGANTGANTVIVLSDG 60
 DB 31 nansfyvsgtlydangpfrmginhanykdattatgiantgntvrvlvsdgs 90
 OY 61 QWTKDHTVRLLSLAEDNHLVAPEVDATGYDSIASLANAVYVTEMRALGKEVT 120
 DB 91 qwtkdhtvrllslaednhlvaapevdagyslaslanavvyvtemrallgkect 150
 OY 121 VLTINANWFSGWEDAMADGYKQAIPLRNAGLWHTLWVDAAGWQFPOSIDHGREVF 180
 DB 151 vltinanevfgswedawadgykqatprlinsglnhtlwnvdaagvqfpgsidhgrevf 210
 OY 161 NADPQNTWFSIHMTETVAGNNSQVFTNIDRVLNDLALVIGFGRHRTNGDVDEATIMS 240

DB 211 naddpqrntwfsihmetyagngasqvrtnidrvlndlalvigeighrtngdvdeatims 270
 OY 241 YSEQRGVGTAWSMKNGNPEWEYIDLSDMAGNNLTANGNTVNGPYGLRETSRLSTVET 300
 DB 271 yseqrvgvgtawsmkngnpeweyldlsdmagnnltangntvngpyglretsrilstvft 330
 RESULT 2
 Y54123
 ID Y54123 standard; Protein; 476 AA.
 Y54123;
 DT 27-MAR-2000 (first entry)
 DE A mannanase-linker-cellulose binding domain fusion protein.
 XX
 XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; *Bacillus* sp. 1633; galactomannan;
 KW 1,4-beta-D-mannosidic linkage; mannan; galactomannan; glucomannan;
 KW galactoglucmannan; cellulosic fibre; synthetic fibre; yarn; fabric;
 KW printing paste; plant material degradation; recycled waste paper;
 KW paper making pulp; guar; locust bean gum; thickener; viscosity;
 KW mannan-containing food; coffee extract; cleaning composition;
 KW machine washing; hard-surface cleaner; dishwashing; oral; dental;
 KW contact lens; body-care composition; fabric softener; oil well drilling;
 KW subterranean formation fracture; cellulose binding domain.
 XX
 OS Synthetic.
 OS *Bacillus* sp.
 OS *Clostridium thermocellum*.
 XX
 XX WO9964619-A2.
 XX
 PD 15-DEC-1999.
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 XX (NOVO) NOVO-NORDISK AS.
 XX
 XX Kaupplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 XX WPI; 2000-105891/09.
 XX
 XX N-PSDB; 245335;
 XX
 PT New mannanses for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 PS Example 4; Page 211-212; 242pp; English.
 XX
 XX The present sequence represents a mannanase-linker-cellulose binding
 CC domain fusion protein. Mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase)
 CC hydrolyses galactomannans. Specifically, mannanses hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
 CC and galactoglucmannans. The mannanase protein, or preparations
 CC containing it, are used to improve properties of cellulosic or

CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC celluloses; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.

XX Sequence 476 AA;

Query Match 99.6%; Score 1624; DB 21; Length 476;

Best Local Similarity 100.0%; Pred. No. 2, 6e-140; Mismatches 0; Gaps 0;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ANSGFTYSGITLYDANGNPFYMRGINHGHANYKDOATTAGIANTGANTVRYVLSGQ 61
 DB 1 ansgfytsgitlydangnpfymrginhghanykddattalegiantgntvryvlsdgq 60
 QY 62 WTKDDIHTVRLISLAEDNHLVAVPEYHDATGYDSISLNRAVDYWIEMRSALIGKEDTV 121
 DB 61 wtkddihvrrlislaednhlvavpevhdatsysalslnravdywlemrallgkedtv 120
 QY 122 IINIANEFSGMEDAMADGKQAIPLRLNAGLNTLAVDAAGWGPQSIHDYGREYFN 181
 DB 121 iinianeefsgmedawadgskqalprlrnaglnhtlmvdaagvqfqsldhdygrefn 180
 QY 182 ADPQNTMFSIHMYEYAGNAGSOYRINIDRYLNODLALVIGFGRHRTNGVDVDAITMSY 241
 DB 181 adpqnmtfshmyeyagnasgyrindrylnodlalvigefgrhrtngvdeaitmsy 240
 QY 242 SEQGVGWLANSWKGNGPEWYELDSNDWAGNNTLVANGNTVNGPYGRSRLSTVTF 300
 DB 241 seqgvglawswkgngpeweyldlsndwagnnltvngntlvngpygrlstrstvt 299

RESULT 3
 ID Y44496 standard; Protein; 468 AA.
 AC Y44496;
 XX 27-MAR-2000 (first entry)
 DE Bacillus agaradhaerens Clone MB594, Mannanase enzyme.
 XX
 KW Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
 KW detergent composition; mild-branched anionic surfactant; washing;
 KW cosmetic stain; food stain.
 XX
 OS Bacillus agaradhaerens.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..31
 FT /label= signal_peptide
 FT Protein 32..468
 FT /label= Mature_Mannanase
 XX
 XX MO9964552-A1.
 XX
 XX 16-DEC-1999.
 PD
 PF 10-JUN-1998; 98MO-US12026.
 XX
 XX 10-JUN-1998; 98MO-US12026.
 PR
 XX 10-JUN-1998; 98MO-US12026.
 XX
 PA (PROC) PROCTER & GAMBLE CO.

XX Bettloll JP, Thoen CAUK;

XX WPI; 2000-116536/10.

DR N-PSDB; 229846.

PT Detergent composition for removing greasy stains such as cosmetics,
 PT food stains and body soils -

XX Disclosure; Page 103-104; 113pp; English.

CC The present sequence is the B. agaradhaerens Clone MB594, alkaline
 CC mannanase enzyme. Clone MB594 is derived from B. agaradhaerens strain
 CC NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging
 CC from 7.5-10.5. It can be used in a detergent composition along with a
 CC mild-branched anionic surfactant. The detergent composition may be used
 CC for washing purposes, soaking/pre-treatment of stained fabric, hard
 CC surface cleaning and for removal of cosmetic and/or food stains. This
 CC composition provides excellent cleaning effect at low temperature.

XX Sequence 468 AA;

Query Match 78.7%; Score 1283; DB 21; Length 468;

Best Local Similarity 77.3%; Pred. No. 3, 5e-109; Mismatches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

QY 2 ANSGFTYSGITLYDANGNPFYMRGINHGHANYKDOATTAGIANTGANTVRYVLSGQ 61
 DB 32 asgfytsgitlydangnpfymrginhghanykdtastalpallegantlrvlsdgq 91
 QY 62 WTKDDIHTVRLISLAEDNHLVAVPEYHDATGYDSISLNRAVDYWIEMRSALIGKEDTV 121
 DB 92 wtkddihvrrlislaednhlvavpevhdatsysalslnravdywlemkallgkedtv 151
 QY 122 IINIANEFSGMEDAMADGKQAIPLRLNAGLNTLAVDAAGWGPQSIHDYGREYFN 181
 DB 152 iinianeefsgmedawadgskqalprlrnaglnhtlmvdaagvqfqsldhdygrefn 211
 QY 182 ADPQNTMFSIHMYEYAGNAGSOYRINIDRYLNODLALVIGFGRHRTNGVDVDAITMSY 241
 DB 212 adpqnmtfshmyeyagnasgyrindrylnodlalvigefgrhrtngvdeaitmsy 271
 QY 242 SEQGVGWLANSWKGNGPEWYELDSNDWAGNNTLVANGNTVNGPYGRSRLSTVTF 300
 DB 272 seqgvglawswkgngpeweyldlsndwagnnltvngntlvngpygrlstrstvt 330

RESULT 4
 ID Y54125 standard; Protein; 468 AA.
 AC Y54125;
 XX 27-MAR-2000 (first entry)
 DE Amino acid sequence of a Bacillus mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX
 XX Synthetic.
 OS Bacillus agaradhaerens.
 OS
 XX
 XX MO9964619-A2.

```

Db 212 adplknmsimhyegadqantvrenidavldqdlavlgcghnhtdgdveditlly 271
QY 242 SEQRGVWGLAMSKNGPWEYLDLSNDWAGNNLTAVGNTIVGPGGLRFTSLRYET 300
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 272 seegtgvlawsvkngstewdyldlsedvaghlltdgwrivngadglgetskpstyft 330

RESULT 5
Y44495
Y44495 standard; Protein; 493 AA.
XX Y44495;
XX
XX 27-MAR-2000 (first entry)
XX
XX Bacillus agaradherens NCIMB 40482, Mannanase enzyme.
DE
XX
XX Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing;
KW cosmetic stain; food stain.
XX
XX Bacillus agaradherens.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..32
XX FT /label= Signal_peptide
XX FT 33..343
XX FT /label= Mature_Mannanase
XX
XX PN WO9964552-A1.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX 10-JUN-1998; 98WO-US12026.
XX
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Bettloli JP, Thoen CAUK;
XX
XX WPI; 2000-116536/10.
XX
XX N-PSDB; 229845.
XX
XX Detergent composition for removing greasy stains such as cosmetics,
XX food stains and body soils
XX
XX Disclosure; Page 102; 113pp; English.
XX
XX The present sequence is the B. agaradherens NCIMB 40482, alkaline
XX mannanase enzyme. Mannanase shows maximum activity at pH ranging
XX from 7.5-10.5. It can be used in a detergent composition along with a
XX mid-branched anionic surfactant. The detergent composition may be used
XX for washing purposes, soaking/pre-treatment of stained fabric, hard
XX surface cleaning and for removal of cosmetic and/or food stains. This
XX composition provides excellent cleaning effect at low temperature.
XX
XX Sequence 493 AA:

Query Match 78.7%; Score 1283; DB 21; Length 493;
Best Local Similarity 77.3%; Pred. No. 3.7e-109;
Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0

QY 2 ANSGFVSGTTLTYDANGNPFVNRKGINHGHWYKKDQATTAIEGIANTGANTVRIYSDGQ 61
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 32 astgfyvngntlydangqpfvmrginhgawykdstaipaiaeqgantlrvlsdsgq 91
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 WTKDDIHTYRNILSLAEDNHLVAPVYHDATGYDLSASINRAVDVYIEMRSALIGKEDTV 121
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 92 wekdidditlrevelaedqkmaavvevhatgrdsrdlnravdyvlemkdallgkcdv 151
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

OY 122 IININEMFGSWEGDAMADGYKQAIPLRLNAGLNTLWDAWGQFPOSTHIDGREYFN 181
 DB 152 IININEMFGSWEGDAMADGYKQAIPLRLNAGLNTLWDAWGQFPOSTHIDGREYFN 211
 OY 182 ADPQNTMFSIMMYEYAGGNSQVNTNIDRYLNODLALVIGFGRHNTGVDDEATIMSY 241
 DB 212 adpLxntmfsImyeyaggdantvrsndrldqdlavlgfgrhntgvddeatllsy 271
 OY 242 SEQRGVWGLAMSGKNGPMEYELDISNDWAGNNLWAMGNTIYNGPYGLRETSRLSTVFT 300
 DB 272 seetgtgylawswkngstewdyldlsewagqhlctwgnrlvthgagdgqetskpsvtft 330
 RESULT 6
 Y54124 ID Y54124 standard; Protein; 493 AA.
 AC Y54124;
 DT 27-MAR-2000 (first entry)
 DE Amino acid sequence of a *Bacillus mannanase* enzyme.
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; celluloseic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 KW *Bacillus agaradhaerens*.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..31
 FT /note= "signal peptide"
 FT Domain 32..344
 FT /note= "catalytic domain"
 FT Domain 345..493
 FT /note= "domain of unknown function"
 XX
 PN W09964619-A2.
 PD 16-DEC-1999.
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX WPI: 2000-105891/09.
 DR N-PSDB: 245337.
 XX
 PT New mannanases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions

XX
 PS Claim 35; Page 213-214; 242pp; English.
 CC
 CC The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The
 CC mannanase hydrolyses galactomannans. Specifically, mannanases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans,
 CC and galactoglucomannans. The mannanase protein, or preparations
 CC containing it, are used to improve properties of cellulosic or
 CC synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics), as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 CC
 SQ Sequence 493 AA;
 Query Match 78.7%; Score 1283; DB 21; Length 493;
 Best Local Similarity 77.3%; Pred. No. 3.7e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;
 OY 2 ANSGRYVSGTTLXYDANGNPFVWRGINHGHWKQDQATAEISANTGANTVRIVSDGQ 61
 DB 32 astgfyvgntlydangpfrvmrghhwykdaastalpaiegantlvivsdgq 91
 OY 62 WTKDDIHTVRNLISLAEEDNHLVAVPEVHDATGYDSIASLNRAVDYIEMRSALICKEDTV 121
 DB 92 wekdldtltreivelaegnkhwavrevndatgdsrdlnravdvwmekdalisgedtv 151
 OY 122 IININEMFGSWEGDAMADGYKQAIPLRLNAGLNTLWDAWGQFPOSTHIDGREYFN 181
 DB 152 IININEMFGSWEGDAMADGYKQAIPLRLNAGLNTLWDAWGQFPOSTHIDGREYFN 211
 OY 182 ADPQNTMFSIMMYEYAGGNSQVNTNIDRYLNODLALVIGFGRHNTGVDDEATIMSY 241
 DB 212 adpLxntmfsImyeyaggdantvrsndrldqdlavlgfgrhntgvddeatllsy 271
 OY 242 SEQRGVWGLAMSGKNGPMEYELDISNDWAGNNLWAMGNTIYNGPYGLRETSRLSTVFT 300
 DB 272 seetgtgylawswkngstewdyldlsewagqhlctwgnrlvthgagdgqetskpsvtft 330
 RESULT 7
 Y54127 ID Y54127 standard; Protein; 331 AA.
 AC Y54127;
 DT 27-MAR-2000 (first entry)
 DE Amino acid sequence of a *Bacillus mannanase* enzyme.
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; celluloseic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 KW *Bacillus* sp.
 OS
 XX
 FH Key Location/Qualifiers

| Query Match | Best Local Similarity | 69.7% | Score 1136 | DB 21 | Length 331 |
|-------------|-------------------------------------------------------------------|---------------|------------|--------|------------|
| Matches 204 | Conservative 47 | Mismatches 46 | Indels 0 | Gaps 0 | |
| 3 | NSGFVSGTGLTDANGNPVVRMGINNGHMYRQDATTATGATNGANVRYIYSPGGOM | 62 | | | |
| Db | 34 hsgfmgntgltdangnpvrmginhghawirgqelactsmgslsqgancitrvlslngqrw | 93 | | | |
| 0Y | 63 TRDHTHTVNNLLSLAEDNHLVAPEVHADATGYDSIASLNRAVDYVTEMKSAALIGKEDTYT | 122 | | | |
| Db | 94 qddndnmvavslslaeqhmqlavlrvhdatsgmfnfdlgaavdyvtemkdvlgqkediyl | 153 | | | |
| 0Y | 123 INIAENWFGSWGMDAMADYKQAIPIRLRNAGLNHTLMVDYDAAGGCPFGPSIHIDYGRVFN | 182 | | | |

| | | |
|--------|-------------------------------------------------------------------------------|-------------------------------------------------------------------|
| Dd | 154 | linineyrgwdggawarygynatrlrlnaglshtclmvaasyqypgavvdyggevlina |
| Oy | 183 | DPOSNTFSTHMYETAYAGNAsQYRINIDRYLNODLAVIGEGFHRTNGDVDEATIMSYs |
| Dd | 214 | dprntfsvlmuyeyaggdantvrrlidsllsqalvalgefshwlydgdvdadctllsyS |
| Oy | 243 | EORGVGMIAWSKCNPEWFEYLDSMDMAGNNLTAMGNTIVNGPGRLRSTSLTVE |
| Dd | 274 | qgrnvglawswngnsegveylalsndfgnrltwgdrlvngpqrqltskrssvf |
| RESULT | 8 | |
| ID | Y54129 | standard; Protein: 369 AA. |
| AC | Y54129; | |
| DT | 27-MAR-2000 | (first entry) |
| XX | | Amino acid sequence of a Bacillus sp. A349 mannanase enzyme. |
| KW | Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase; | |
| KW | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; | |
| KW | mannan; galactotriannan; glucotriannan; galactoglucotriannan; cellulose fibre | |
| KW | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; | |
| KW | plant material degradation; recycled waste paper; paper making pulp; | |
| KW | guar; locust bean gum; mannan-containing food; coffee extract; | |
| KW | cleaning composition; machine washing; hard-surface cleaner; | |
| KW | dishwashing; oral; dental; contact lens; body-care composition; | |
| KW | fabric softener; oil well drilling; subterranean formation fracture. | |
| OS | Bacillus sp. | |
| XX | | |
| XX | Key | Location/Qualifiers |
| FT | Peptide | 1..25 |
| FT | | /note= "signal peptide" |
| FT | Domain | 26..67 |
| FT | | /note= "N-terminal domain of unknown function" |
| FT | Domain | 68..369 |
| FT | | /note= "catalytic domain" |
| PN | WO9964619-AZ. | |
| PD | 16-DEC-1999. | |
| PF | 10-JUN-1999; | 99WC-DX00314. |
| PR | 10-JUN-1998; | 98US-0111256. |
| PR | 20-OCT-1998; | 98DK-0001340. |
| PR | 20-OCT-1998; | 98DK-0001341. |
| PR | 28-OCT-1998; | 98US-0105970. |
| PR | 28-OCT-1998; | 98US-0106054. |
| PR | 23-DEC-1998; | 98DK-0001725. |
| PR | 05-MAR-1999; | 99DK-0000306. |
| PR | 05-MAR-1999; | 99DK-0000307. |
| PR | 05-MAR-1999; | 99DK-0000308. |
| PR | 05-MAR-1999; | 99DK-0000309. |
| PR | 09-MAR-1999; | 99US-0123843. |
| PR | 10-MAR-1999; | 99US-0123823. |
| PR | 10-MAR-1999; | 99US-0123641. |
| PR | 11-MAR-1999; | 99US-0123642. |
| PA | (NOVO) NOVO-NORDISK AS. | |
| Pt | Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME; | |
| DR | N-PSDB; Z45342. | |
| XX | | New mannases for treatment of textiles, plant material and coffee |
| XX | | extract, and in cleaning compositions - |

Db 147 vlinanewyawsrsvaasayagalpirlsaglahlilvdaagvgypasihergadvf 206
 Oy 181 NADPORNTEFSTHMYEYAGNAGSVYRINIDRYLNQDLALVIGFEHRTNGVDDEATIMS 240
 Db 207 asdpknctmfstlmeyegadatrsvsenldgylaenlavigefghrhdgdvdedalla 266
 Oy 241 YSEORGVGLAWSKNGKNGPEWETDLSDMAGNNITLNGNTIVNCPYGLRETSRL 295
 Db 267 ylaervgvylawswygnsgvgyldltgpgsg-pltswgerlvygmglykvidhl 320

RESULT 10
 ID Y54133 standard; Protein; 188 AA.
 AC Y54133;
 DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a partial *Bacillus* sp. mannanase enzyme.
 XX
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS *Bacillus* sp.
 XX
 PN WO964619-A2.
 PD 16-DEC-1999.
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 DR WPI; 2000-105891/09.
 DR N-PSDB; 245346.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions
 PT
 XX
 XX Claim 35; Page 229-230; 242pp; English.
 XX
 CC The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 XX
 SQ Sequence 188 AA;
 Query Match 32.0%; Score 522; DB 21; Length 188;
 Best Local Similarity 59.9%; Pred. No. 3e-40;
 Matches 97; Conservative 20; Mismatches 45; Indels 0; Gaps 0;

Oy 1 NANGFYVSGTITVDANCPVFMGNGHAWYKQATFATEGIANTGANTVRIVLSDGG 60
 Db 27 saqgfhvkgtellidkngdpyrmgyvnhgshwtkqdlleaipaetcyanvrvlvsngq 86
 Oy 61 QWTKDITVNTLSIAEDNHLVAPVEYDATGYDSIASLNRAVDYWIEMRSALIGREDT 120
 Db 87 qwekddasaelarylaetctyltvtlevhdatsgdndpddlkavdyvlemadvlykgtedr 146
 Oy 121 VIINTANFEGSWEGDAMADGKQALPRLNAGLNTMTMDVA 162
 Db 147 vlinanewyawsrsvdhwakayagalpirlsaglahlilvda 188

RESULT 11
 ID Y54130 standard; Protein; 305 AA.
 AC Y54130;
 DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a partial *Bacillus* sp. mannanase enzyme.
 XX
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS *Bacillus* sp.
 XX
 PN WO964619-A2.
 PD 16-DEC-1999.
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.

PR 11-MAR-1999; 99US-0123642.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI
PI WPI; 2000-105891/09.
DR N-PSDB; 245343.
XX
XX New mannases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
XX
XX Disclosure; Page 225-226; 242pp; English.
XX
XX The present sequence represents a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaner, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.
XX
XX Sequence 305 AA;

Query Match 22.6%; Score 369; DB 21; Length 305;
Best Local Similarity 45.3%; Pred. No. 5.4e-26;
Matches 73; Conservative 32; Mismatches 46; Indels 10; Gaps 4;
XX
XX 146 IPRLRNGLNLTMDVADGAGQFPOSIDHYEFNADPQNTFTSIHMETYAGNASQV 205
DB 1 IRLRLAGNLTIVDASGYSQNSPKAYNEVLNDPGRNMFSLHMYG-SWNGSRL 59
XX
XX 206 RINIDRYNDLALVIGEFCHRTNGD-----VDENTISYSEQRGVGLANSMKNGP 259
DB 60 GSELIGALIKDLGLVMIIGEFYDNGNMLGSQVNAQELMGQAKYGYMPYWTGDA 119
XX
XX 260 EWEYLIDL-SNDMAGNNLTAGNTIVNGPYGLRERSRLSTVF 299
DB 120 ANSWLMDTNDW--qtltsvgnlvvngtnglratsvpavtf 158

RESULT 12
Y54J31
ID Y54J31 standard; Protein: 132 AA.
XX
XX Y54J31;
AC
XX 27-MAR-2000 (first entry)
DT
XX
DE Amino acid sequence of a partial Bacillus mannanase enzyme.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW mannan; galactomannan; glucomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactoglucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture.
XX

OS Bacillus clausii.
XX
XX W09964619-A2.
XX
XX 16-DEC-1999.
XX
XX 10-JUN-1999; 99WO-DK00314.
PE
XX 10-JUN-1998; 98US-0111256.
XX 20-OCT-1998; 98DK-0001340.
XX 20-OCT-1998; 98DK-0001341.
XX 28-OCT-1998; 98US-0105970.
XX 28-OCT-1998; 98US-0106054.
XX 23-DEC-1998; 98DK-0001725.
XX 03-MAR-1999; 99DK-0000306.
XX 03-MAR-1999; 99DK-0000307.
XX 05-MAR-1999; 99DK-0000308.
XX 05-MAR-1999; 99DK-0000309.
XX 09-MAR-1999; 99US-0123543.
XX 10-MAR-1999; 99US-0123523.
XX 10-MAR-1999; 99US-0123641.
XX 11-MAR-1999; 99US-0123642.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
PI
PI WPI; 2000-105891/09.
DR N-PSDB; 245344.
XX
XX New mannases for treatment of textiles, plant material and coffee
PT extract, and in cleaning compositions -
XX
XX Disclosure; Page 227; 242pp; English.
XX
XX The present sequence represents a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.
XX
XX Sequence 132 AA;

Query Match 19.5%; Score 318; DB 21; Length 132;
Best Local Similarity 55.6%; Pred. No. 7.8e-22;
Matches 60; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
XX
XX 2 ANSGFYSGTTLTDANGNPFVKGINHGAMRYDQATTAIEGANTGANTVRIYVLDGQ 61
DB 23 AGTGYTGTCGLIDGPNPYVMRGVHGHFWFGDIDTALPATAATGANLTVRIYVLSNG 82
XX
XX 62 WTDGDDIHTVNRNLISLADNHLVAPREHDTAGTDSIASLNRADYWE 109
DB 83 WERTDAVEERVLAVEEGLTAVLEHATGSDGDPNDITAVEYSE 130
XX
XX RESULT 13
W12378
ID W12378 standard; Protein: 410 AA.

```

XX AC W12378;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 1 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. N4 (ATCC 21833).
XX FH Key
XX FT 1.26 Location/Qualifiers
XX FT /label= Sig.peptide
XX FT /note= "hybrid between P300 (aa1-12) and
XX FT CelB (aa13-26) signal peptides"
XX FT Protein
XX FT 27..410
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN W09706181-A1.
XX PD 20-FEB-1997.
XX PF 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENK ) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB; T63226.
XX FT System for increased expression of cellulase and xylanase in
XX FT Bacillus - contains gene under control of elements from B.
XX FT licheniformis alkaline protease gene
XX PS Disclosure; Fig 7,7A,7B; 37pp; English.
XX CC The polypeptide product (W12378) of P300-CelB fusion construct 1
XX CC (T63226) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also
XX CC W12379-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 410 AA;

Query Match 12.4%; Score 202; DB 18; Length 410;
Best Local Similarity 25.0%; Pred. No. 1.4e-10;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

OY 8 VSGTLYDANGNFVARGI-NHGHWYKDOATTAIGIANFGANWVRIYLSDGQGWTKDD 66
DB 41 lmgelvdnrgpvdqkmgshglqyq-----qfnysasmwlrdd 82
OY 67 -----IHT-----VRNLSIAEDNHLVWVPEWHDANGYSISLNRA 103
DB 83 wgtlvraamytesaggyiedpsvkekvkeaveaaidglgylldwllldndpnlykeea 142
OY 104 VDWIMRSLILIKEDPTIININANWFGS---WEGD--AMADGYKQAIPLRNAGLHTL 158
DB 143 kdfidem-selygdynvlyelanepngsdvtwngdkpyae---evipvlnndpnll 198

```

```

OY 159 WYDAAGMGCFPOSIHDIKREVFNADPQNTMFSTHMTAYAGNSAOYRNTDRLNODLA 218
DB 199 lv---gtgtwsqdvh-headnqltdp--nymyafhf--yagthynqlrtdyvaldgaa 250
OY 219 LVYGEFGHRTNGD-----VDATI-MSYSEORGWGLAWS 253
DB 251 lrvsewgtseatgdgvtldeaqwldmdernlswanws 290

RESULT 14
W12379
ID W12379 standard; Protein; 411 AA.
XX AC W12379;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 2 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. N4 (ATCC 21833).
XX FH Key
XX FT 1.26 Location/Qualifiers
XX FT /label= Sig.peptide
XX FT /note= "hybrid between P300 (aa1-5) and
XX FT CelB (aa8-26) signal peptides"
XX FT Protein
XX FT 27..411
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN W09706181-A1.
XX PD 20-FEB-1997.
XX PF 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENK ) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB; T63227.
XX FT System for increased expression of cellulase and xylanase in
XX FT Bacillus - contains gene under control of elements from B.
XX FT licheniformis alkaline protease gene
XX PS Disclosure; Fig 8,8A,8B; 37pp; English.
XX CC The polypeptide product (W12379) of P300-CelB fusion construct 2
XX CC (T63227) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. N4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also W12378,
XX CC W12380-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 411 AA;

Query Match 12.4%; Score 202; DB 18; Length 411;
Best Local Similarity 25.0%; Pred. No. 1.4e-10;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:02:41 ; Search time 156.76 Seconds
(without alignments)
32.077 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630

Sequence: 1 NANSGRVYSCTILYDANGNP.....TIVNGPYGLRHSRLSTVPT 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA: *
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 201 | 12.3 | 400 | 2 | US-08-713-298B-2 |
| 2 | 201 | 12.3 | 400 | 2 | US-08-870-180B-2 |
| 3 | 201 | 12.3 | 400 | 3 | US-08-814-052-4 |
| 4 | 201 | 12.3 | 400 | 3 | US-08-812-829-4 |
| 5 | 201 | 12.3 | 462 | 2 | US-08-870-180B-13 |
| 6 | 191 | 11.7 | 467 | 2 | US-08-727-548-2 |
| 7 | 171.5 | 10.5 | 430 | 2 | US-08-924-440-2 |
| 8 | 142 | 8.7 | 358 | 1 | US-08-604-913B-11 |
| 9 | 142 | 8.7 | 521 | 1 | US-08-276-213-3 |
| 10 | 135 | 8.3 | 531 | 2 | US-09-033-537A-1 |
| 11 | 110 | 6.7 | 531 | 2 | US-07-862-388B-7 |
| 12 | 100 | 6.1 | 429 | 1 | US-08-745-977-4 |
| 13 | 100 | 6.1 | 429 | 1 | US-09-040-699A-4 |
| 14 | 99.5 | 6.1 | 317 | 2 | US-09-066-075-2 |
| 15 | 99.5 | 6.1 | 317 | 2 | US-08-518-615A-2 |
| 16 | 99.5 | 6.1 | 317 | 3 | US-08-951-889-2 |
| 17 | 97.5 | 6.0 | 574 | 3 | US-08-733-433-1 |
| 18 | 95.5 | 5.9 | 735 | 2 | US-08-313-185-8 |
| 19 | 95.5 | 5.9 | 735 | 2 | US-08-459-499-9 |
| 20 | 94 | 5.8 | 398 | 2 | US-08-853-659A-45 |
| 21 | 93.5 | 5.7 | 566 | 2 | US-07-863-588B-4 |
| 22 | 90.5 | 5.6 | 735 | 2 | US-08-459-499-12 |
| 23 | 88 | 5.4 | 531 | 1 | US-08-531-601-1 |
| 24 | 88 | 5.4 | 531 | 2 | US-08-859-032-1 |
| 25 | 88 | 5.4 | 535 | 1 | US-08-737-597-10 |
| 26 | 86 | 5.3 | 461 | 1 | US-08-672-571A-3 |
| 27 | 86 | 5.3 | 490 | 1 | US-08-672-571A-1 |
| 28 | 85.5 | 5.2 | 385 | 2 | US-08-387-942C-26 |

| | | | | | | |
|----|------|-----|------|---|-------------------|-------------------|
| 29 | 85.5 | 5.2 | 997 | 2 | US-08-387-942C-4 | Sequence 4, Appl1 |
| 30 | 85.5 | 5.2 | 1248 | 2 | US-08-348-353-17 | Sequence 17, Appl |
| 31 | 85.5 | 5.2 | 1248 | 2 | US-08-465-965-17 | Sequence 17, Appl |
| 32 | 85.5 | 5.2 | 1248 | 3 | US-08-465-966-17 | Sequence 17, Appl |
| 33 | 85 | 5.2 | 881 | 1 | US-08-333-901-1 | Sequence 1, Appl1 |
| 34 | 85 | 5.2 | 881 | 1 | US-08-456-582-1 | Sequence 1, Appl1 |
| 35 | 85 | 5.2 | 881 | 2 | US-08-898-789-1 | Sequence 1, Appl1 |
| 36 | 85 | 5.2 | 881 | 3 | US-09-039-555B-16 | Sequence 16, Appl |
| 37 | 85 | 5.2 | 4544 | 1 | US-08-469-486-52 | Sequence 52, Appl |
| 38 | 85 | 5.2 | 4544 | 2 | US-08-469-658-52 | Sequence 52, Appl |
| 39 | 84.5 | 5.2 | 2628 | 2 | US-08-570-311-14 | Sequence 14, Appl |
| 40 | 84 | 5.2 | 3724 | 2 | US-08-804-227C-10 | Sequence 10, Appl |
| 41 | 84 | 5.2 | 3724 | 2 | US-08-804-198-4 | Sequence 4, Appl1 |
| 42 | 83 | 5.1 | 514 | 3 | US-08-600-656-3 | Sequence 3, Appl1 |
| 43 | 83 | 5.1 | 531 | 3 | US-08-688-988-39 | Sequence 39, Appl |
| 44 | 82 | 5.0 | 455 | 2 | US-08-870-827-3 | Sequence 3, Appl1 |
| 45 | 81.5 | 5.0 | 557 | 2 | US-08-793-229-33 | Sequence 33, Appl |

ALIGNMENTS

RESULT 1
US-08-713-298B-2
Sequence 2, Application US/08713298B
Patent No. 5922586
GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dammann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Sch Ielh, Martin
APPLICANT: J rgensen, Per
TITLE OF INVENTION: DNA Constructs and Methods of Producing
TITLE OF INVENTION: Cellulolytic Enzymes
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 59225860 No. 5922586dsk of No. 5922586th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,298B
FILING DATE: 13-SEPT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3794, 424-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 867-0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-298B-2
Query Match 12.3%, Score 201; DB 2; Length 400;
Best Local Similarity 26.6%, Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

```

0Y      8 VSGTLLYANGNPEFMEFRI--HNGAWY-----KQOATTALEGJANTGANVRI 54
      40 ISNGELVNERGEQVQJLKMSSHGJLQWIGQFNYTSMKMLR-----DMGJINFERA 89
0Y      55 VL--SDGOWTKDDI--HTVRNLISLAEDELNLVAVPEVHADATGYDSIASLNBAVDYIEMR 111
0Y      90 AMYTSSGGYIDDPYKVEYKKAVERAAILDLITYIIDHMLISDNDPNYKKEADFEDEM 148
0Y      112 SALLKEPTVLININEMFEGMEGDAMADGK---CAIPLRAGNLHTLMDVADAGWQ 167
0Y      149 SELYDDYNNVIELANENGSS--DYTWGNQKPLAEVYIPIIRNNDRNNIIV---GTGT 203
0Y      168 FPOSIHIDYGREVENADPORNTMESIHMYEYAGGASOVRTNIDEVLNODIALYIGEGR 227
0Y      204 WSOVDF--HAALNQLADP--NMYAFHF--YAGTGQMLRDQVDYALDQGAIFSEWGT 258
0Y      228 HTNGD---VDEART--MSYSOEGOGYLAWS 253
0Y      259 AATGDDGYFIDELQOYWIDFMDERNISMANWS 289

```

RESULT 2
 US-08-870-180B-2
 : Sequence 2, Application US/08870180B
 : Patent No. 5945327
 : GENERAL INFORMATION:
 : APPLICANT: Outtrup, Helle
 : APPLICANT: Dammann, Claus
 : APPLICANT: Olsen, Arne
 : APPLICANT: Bisg rd-Frantzen, Henrik
 : APPLICANT: Sch lein, Martin
 : APPLICANT: J rgensen, Pjet
 : APPLICANT: Bjoernvad, Mads
 : TITLE OF INVENTION: DNA Constructs and Methods of Producing
 : TITLE OF INVENTION: Cellulytic Enzymes
 : NUMBER OF SEQUENCES: 14
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 59453270 No. 5945327/disk of No. 5945327th America, Inc.
 : STREET: 405 Lexington Avenue, Suite 6400
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10174-6401
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25 (EPC)
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/870,180B
 : FILING DATE: 6-June-1997
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Rozek, Carol E.
 : REGISTRATION NUMBER: 36, 993
 : REFERENCE/DOCKET NUMBER: 3794.434-US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 212 867 0123
 : TELEFAX: 212 867 0298
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 400 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-870-180B-2

Query Match 12.3% Score 201 DB 2 Length 400;
Best Local Similarity 26.6% Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;
0Y 8 VSGTLLYPDANGNPFVWRCI-NHGHWY-----KQDATTALBETIANGTANTYRI 54

```

Db      40  ISNCELVNEREQVOLGMSHGIOMTGOEFVNESMKLAD-----DWGINFRA  89
      41  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      55  VL--SDGOWTKRDI-RHYRRLISLADNHLVAVPEVHDATGYDSIASLRADVYTEMR  111
      56  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      90  AMYTSGGYIDDPVKRKKEVNEAALIDDIYIIDHILSDNDPNRYKREADFPEM-  148
      91  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     112  SALLGKEDYIINIANWEGSWBEDDAADGK-----QAIRLRNAGINHLMDAAGWQ  167
      113  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     149  SELGIDYPNAYEIEIANEPNGS--DYTGNOIKRPAEEVITPIIRNNDPNNIIIV--GTGT  203
      150  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     168  FPOSIDHYGRVFNADQORNTMFESINHXEAGNAGSQRINIDIRVLMQDLALYIGEGHR  227
      169  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     204  WSOQVH-HAADNQLADB-VNMYTFHR-YAGTGHQNLBDQVYALDOGAIFVSEWGS  258
      205  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     228  HTNGD----VDEATI-MSYSFORGVGLAMS  253
      229  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     259  AATGDGCVFLDEAQWVLTIDPFMDERLSTAMNS  289
      300  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 3
US-08-814-052-4
: Sequence 4, Application US/08814052
: Patent No. 6015783
: GENERAL INFORMATION:
: APPLICANT: von der Osten, Claus
: APPLICANT: Cherry, Joel R.
: APPLICANT: Bjornvad, Mads E.
: APPLICANT: Vind, Jesper
: APPLICANT: Rasmussen, Michael Dolberg
: TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
: TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 60157830 No. 6015783disk of No. 6015783th America, Inc.
: STREET: 405 Lexington Avenue, Suite 6400
: City: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/814.052
: FILING DATE: 06-MAR-1997
: CLASSIFICATION: 510
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambidis, Elias J
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 4684.204-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 400 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-814-052-4

```

```
Query Match      12.3% Score 201; DB 3; Length 400;
Best Local Similarity 26.6%; Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14.
```

Db 40 ISNGLVNERBQVOYLKGMSSHG:LOWYGOFYVYESMKMLRD-----DWGINVRA 89

QY 55 VL--SDGOWTKDDI-HIVRNLI:SLAEDNHLVAVPEVDATGYDSIASLNRADVWIEKR 111

Db 90 AMYISSGGYIDDP:SVKEKYEKAEVNAIDLDIYIIDWHILSNDPNITKEEKDFDEK- 148

QY 112 SALIGKEPTVLIINANEMFGSWEGDAMADGYK---QAIPLRNAGLHHTLAVDAAGMQ 167

Db 149 SELGDIYRNVIYEI:ANENGS--DYTWGNQIKPYAEVPIIRNNDPNITIV---GTGT 203

QY 168 FPOSIDHYGEVFNADPQRTNFTSIHTEYAGNASQVRTNIDRYLNODLALVGEFGHR 227

Db 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLRDQVDYALDQGAALFVSEWGT 258

QY 228 HTNGD---VDEPAT-MSYSEORGYGWLAMS 253

Db 259 AATGDGCVFLDEAQWIDFMDERNLSMANWS 289

RESULT 4

US-08-812-829-4

; Sequence 4, Application US/08812829

; Patent No. 6017751

; GENERAL INFORMATION:

; APPLICANT: von der Osten, Claus

; APPLICANT: Bjornvad, Mads E.

; APPLICANT: Vind, Jesper

; APPLICANT: Rasmussen, Michael Dolberg

; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6017751st No. 6017751st of No. 6017751st America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/812,829

; FILING DATE: 06-MAR-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Landlits, Elias J

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 4690, 204-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 400 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-812-829-4

Query Match 12.3%; Score 201; DB 3; Length 400;

Best Local Similarity 26.6%; Pred. No. 2e-11;

Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLLYDANGNPFVWRCI-NHGHAWY-----KDOATTAIGCIANTGANTVRI 54

Db 40 ISNGLVNERBQVOYLKGMSSHG:LOWYGOFYVYESMKMLRD-----DWGINVRA 89

QY 55 VL--SDGOWTKDDI-HIVRNLI:SLAEDNHLVAVPEVDATGYDSIASLNRADVWIEKR 111

Db 90 AMYISSGGYIDDP:SVKEKYEKAEVNAIDLDIYIIDWHILSNDPNITKEEKDFDEK- 148

QY 112 SALIGKEPTVLIINANEMFGSWEGDAMADGYK---QAIPLRNAGLHHTLAVDAAGMQ 167

Db 149 SELGDIYRNVIYEI:ANENGS--DYTWGNQIKPYAEVPIIRNNDPNITIV---GTGT 203

QY 168 FPOSIDHYGEVFNADPQRTNFTSIHTEYAGNASQVRTNIDRYLNODLALVGEFGHR 227

Db 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLRDQVDYALDQGAALFVSEWGT 258

QY 228 HTNGD---VDEPAT-MSYSEORGYGWLAMS 253

Db 259 AATGDGCVFLDEAQWIDFMDERNLSMANWS 289

RESULT 5

US-08-870-180B-13

; Sequence 13, Application US/08870180B

; Patent No. 5945327

; GENERAL INFORMATION:

; APPLICANT: Outtrup, Helle

; APPLICANT: Dammann, Claus

; APPLICANT: Olsen, Arne

; APPLICANT: Bisg, rd-Frantzen, Henrik

; APPLICANT: Sch Jeln, Martin

; APPLICANT: J rgensen, Per

; APPLICANT: Bjornvad, Mads

; TITLE OF INVENTION: DNA Constructs and Methods of Producing

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5945327st No. 5945327st of No. 5945327th America, Inc.

; STREET: 405 Lexington Avenue, Suite 6400

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/870,180B

; FILING DATE: 6-June-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Rozek, Carol E.

; REGISTRATION NUMBER: 36,993

; REFERENCE/DOCKET NUMBER: 3794, 434-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 462 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-870-180B-13

Query Match 12.3%; Score 201; DB 2; Length 462;

Best Local Similarity 26.6%; Pred. No. 2.5e-11;

Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLLYDANGNPFVWRCI-NHGHAWY-----KDOATTAIGCIANTGANTVRI 54

Db 40 ISNGLVNERBQVOYLKGMSSHG:LOWYGOFYVYESMKMLRD-----DWGINVRA 89

QY 55 VL--SDGOWTKDDI-HIVRNLI:SLAEDNHLVAVPEVDATGYDSIASLNRADVWIEKR 111

DB 90 AYTSSGGIIDDPSVKEKKEAVEAALDDIYIIDMHLSDNDRIYKEAKDFDEK- 148
112 SALLIKEDTVIINIANEMFSGEGMADGGR---QAIPRLNAGLNTLMDAAGWGQ 167
DB 149 SELYGYPNVIYEIANEPNGS--DVTWGNQIKPYAEVPIRNDPNIIIV---GTGT 203
OY 168 FPOSIDHGREVFNDPQNTMFSIMYEXAGNASQVNTNDRYLNDLAVIGFGRH 227
DB 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLDDQVYDALDGAALFVSEWGT 258
OY 228 HTNGD---VDEATI-MSYSEORGVGWLAMS 253
DB 259 AATGDGVFLDEAQVWIDPMDERNLSMANWS 289

RESULT 6
US-08-727-548-2
Sequence 2, Application US/08727548
Patent No. 5856165
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
TITLE OF INVENTION: Producing Same
NUMBER OF SEQUENCES: 2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,548
FILING DATE:
CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-727-548-2

Query Match 11.7%; Score 191; DB 2; Length 467;
Best Local Similarity 24.6%; Pred. No. 2.3e-10;
Matches 69; Conservative 49; Mismatches 98; Indels 64; Gaps 14;
OY 8 VSGTLLYDANGNPFYMGCI-NHGAWYKDQATTAEGIANGTGANTVRIYLSGGQWTRD 66
DB 40 ISNGELVNERGEQVQLKGMSSHGLQWYG-----QFVNTESMKWLDD 81
OY 67 -----HT-----VKNLSLADNHLVAVPEYHDATGYDSTASLRA 103
DB 82 WGITVFRAMNTSSGGYIIDDPSVKEKKEVEYAAIDIGIYIIDMHLSDNDPNYKEEA 141
OY 104 VDYIEMTSALIGKEDTVIINIANEMFSGS---WEGD--AMADGKQALPRLNAGLNTL 158
DB 142 KDFDEM-SELYGYPNVIYEIANEPNGSDVTWGNQIKPYAE---EVLPVIRNDPNIIIV 197
OY 159 AYDAAGWQFPOSIDHGREVFNDPQNTMFSIMYEXAGNASQVNTNDRYLNDLAVIGFGRH 218
DB 198 IV---GTGTWSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLDDQVYDALDGA 249
OY 219 LVIGFGRHTNGD---VDEATI-MSYSEORGVGWLAMS 253
DB 250 IFVSEWGTSAATGDGVFLDEAQVWIDPMDERNLSMANWS 289

RESULT 7
US-08-924-440-2
Sequence 2, Application US/08924440
Patent No. 5871550
GENERAL INFORMATION:
APPLICANT: Thomas, Steven R.

APPLICANT: Frits et al.
TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc.
STREET: 925 Gate Mall Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,440
FILING DATE: August 27, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC388
TELEPHONE: (650) 846-7555
TELEFAX: (650) 845-6504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-924-440-2

Query Match 10.5%; Score 171.5; DB 2; Length 430;
Best Local Similarity 24.4%; Pred. No. 1.3e-08;
Matches 77; Conservative 49; Mismatches 122; Indels 67; Gaps 19;
OY 8 VSGTLLYDANGNPFYMGCI-NHGAWYKDQAT-TAIEGIA-NTGANTVRI---VLSGGQ 61
DB 137 VCGIOLDDEHGNPVQLKGMSSHGLQWYDHCITDSSDLAATDKMADIRLSMTIQEGBYE 196
OY 62 WT-----KDDITVKNLSLADNHLVAVPEYHDATGYDSTASLNRADVYIEMTSALIGK 117
DB 197 TNPGRFTDRMH---QLIDMATARGLYIVDWHILTPDPHYNDRAKTFEAIQRHASK 253
OY 118 EDTYIINIANEMFSGEGMAD--GY-KQALPRLNAGLNTLMDAAGWGQFPOSIDH 174
DB 254 TN-VLYETANE---PNGVSAKSTKSTAEVYIPVIRKRDPSVITVTRGWSLIGV-EG 307
OY 175 YGREVFNDP--QNTMFSIMYEXAGNASQVNTNDRYLNDLAVIGFGRHT 229
DB 308 SGRAELIANPYNANINMYAFHF-----ASHADNVLNLRSELSFPVYVTEFGIETY 361
OY 230 NGD-----VDEATIMSYSORGVGWLAMSK-----GNGPEMEYIDL 267
DB 362 TGDDANDFOVMDR--YIDLMAERKIGWTKWYSDFRSGAVFGPGTASGGP----- 411
OY 268 NDWAGNULNAGNTI 282
DB 412 --WGSSTLAKSGQWY 424

RESULT 8
US-08-604-913B-11
Sequence 11, Application US/08604913B
Patent No. 5712142
GENERAL INFORMATION:
APPLICANT: Adey, William S.
APPLICANT: Thomas, Steven R.

APPLICANT: Hammel, Michael E.
APPLICANT: Baker, John O.
APPLICANT: Chou, Yat-Chen
TITLE OF INVENTION: METHOD FOR INCREASING
TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-070-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7499
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: E1-CAT
US-08-604-913B-11

Query Match 8.7%; Score 142; DB 1; Length 358;
Best Local Similarity 23.1%; Pred. No. 7.8e-06;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

2 ANSGF-YVSGTILYDANGNPFVNRGIN-----HGAMNYDQATTAEGIANGA 49
1 AGGWTWHSGRRIIDANNVPRIAGINWGFETCNYVHG-IMSRY-RSMIDQIKSLGY 58
50 NTVRIYLDGCGWTRDDI---HTVRNLISLADNHLVAVPEVDAATGYDSIASLNRADV 106
59 NTRIRLPYS-----DILKPGTMRNSINIFYGMN-----QDLQGLTSLQYMKXIVAY 103
107 -----WIEMRSALT-----GKEDVTIINIANEM 129
104 AGQIGRLIILDRHRPDCSGQSALMTYSSVSSEATWISDLOALQRYKGNPTVVGFDLHNE- 162
130 FGSWGDAMADG-----YKQALPRLRNA--GLNHTLMV-----DAAGW----- 165
163 --PHDPACWGCDDPSIDWRLAERAGNNAVLAVPNLLITVEGVSINGDSYMWGGLQGA 220
166 GQFP-----OSIHGYGREVENADPQRTMFSIHMYEYAGNASQV-RTNIDRVLN 214
221 GQYPVVLANVPLNLYVSAHDYATSYV---PQ--TWFSDDPFP---NNMPEIWMKNNGIYLFN 272
215 QDLALV-IGEFERH--RHNGVDYDEATIMSY---SEQRGVG---WLAWSKRGNG----- 258
273 QAIAPVWGEFETTLQSTTDQTMATLTLYQLRPTAQYGDASFQWTFWSWNPDSGDTGIL 332
259 -PEWEYILDSND 269

Db 333 KDDMQVTVTVKD 344

RESULT 9
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Hammel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 8.7%; Score 142; DB 1; Length 521;
Best Local Similarity 23.1%; Pred. No. 1.3e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

2 ANSGF-YVSGTILYDANGNPFVNRGIN-----HGAMNYDQATTAEGIANGA 49
1 AGGWTWHSGRRIIDANNVPRIAGINWGFETCNYVHG-IMSRY-RSMIDQIKSLGY 58
50 NTVRIYLDGCGWTRDDI---HTVRNLISLADNHLVAVPEVDAATGYDSIASLNRADV 106
59 NTRIRLPYS-----DILKPGTMRNSINIFYGMN-----QDLQGLTSLQYMKXIVAY 103
107 -----WIEMRSALT-----GKEDVTIINIANEM 129
104 AGQIGRLIILDRHRPDCSGQSALMTYSSVSSEATWISDLOALQRYKGNPTVVGFDLHNE- 162
130 FGSWGDAMADG-----YKQALPRLRNA--GLNHTLMV-----DAAGW----- 165
163 --PHDPACWGCDDPSIDWRLAERAGNNAVLAVPNLLITVEGVSINGDSYMWGGLQGA 220
166 GQFP-----OSIHGYGREVENADPQRTMFSIHMYEYAGNASQV-RTNIDRVLN 214
221 GQYPVVLANVPLNLYVSAHDYATSYV---PQ--TWFSDDPFP---NNMPEIWMKNNGIYLFN 272

QY 215 ODLALV-IEEFGH--RHNGDVEATIMSY--SEQRGVG--WLAWSKNG----- 258
DB 273 ONIAPWLEEFGLTOSTDQWTKLTVOLRPTAQYADSEFQWTSWNPNSGDTGIL 332
QY 259 -PEWEYIDLSD 269
DB 333 KDDMOTVDTYVKD 344

RESULT 10

US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fitch, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-033-537A-1

Query Match 8.3%; Score 135; DB 2; Length 551;
Best Local Similarity 20.9%; Pred. No. 6.8e-05;
Matches 81; Conservative 49; Mismatches 133; Indels 124; Gaps 19;

QY 8 VSGTLLYDANGNPFVARGI-NHGHANYKQDATTAGIANTGANTVRIYISDGGQWTKDD 66
DB 11 VQGNQLVGGSGAVQLVQKMSHGLQWYGN-----FVKKSSIQWWRDN 52
QY 67 --HTYRNLISLADNHLVAVPEVHDATGYDTSASLAKRA-----VDYWI----- 108
DB 53 WGINVRAALYTAEDGY-TIDPSVKKNVKEAVASIDLGILYIITDHIISDGNPNYTKAO 111
QY 109 -----EMRSALICEKEDTVIININANWFGSWEGDAMAD--GY-KQATPRLNAGLNHTIMV 160

DB 112 SKAFIOENATLYGNPNPIYELANPNGN---VSNADKSYAEVITAIRADPGGVIV 168
QY 161 DAGWGQFPOSJHDYGREVNADPQRTMFSIHMTYRAGNAGNAGVNTNIDRLNODLALV 220
DB 169 GSPYTSQ-----DIHLADNPVSHSNVYKALHF--YSGTHGQFLRDRITYAMNKGALIF 220
QY 221 IGEFGHRTNGD-----VDEATIMSYSEQRGVGWLAWS-----WKNGPEWEYL 264
DB 221 VTEMGTSDASGNGGPFYFQSKEMIDFLNARKISWVNSLADRVETSAALMPASPTGWT 280
QY 265 DL-----SNDM-----AGN-----NTLAW-GN----- 280
DB 281 DAGLSESGRWYRDOIRQATGGGSGNPAPATNLSATAGNAGVSLTNVAGNASTYTVK 340
QY 281 --TYNGPY-----GIRETSRLSTVFT 300
DB 341 RATTSGPYTNVATGVTATSYNTGIL 367

RESULT 11

US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC Compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-7

Query Match 6.7%; Score 110; DB 2; Length 531;
Best Local Similarity 19.6%; Pred. No. 0.016;
Matches 54; Conservative 45; Mismatches 118; Indels 58; Gaps 10;

QY 8 VSGTLLYDANGNPFVARGI-NHGHANYKQDATTAGIANTGANTVRIYISDGGQWTKDD 66

DB 44 VQGNOLVGGSGAVALVGGSSHGLQATGN-----FNKSSLOMARDN 85
QY 67 -IHTVRNLISAEENHLVAPEVHDATGYDSIASLNRA----VDYWI----- 108
DB 86 WGINVRAMAYTSEDCY-ITDPSVKNKVEAVQASIDLALVYIDHILSDGNPTTKQA 144
QY 109 -----EMRALLGKEDTVIINIANEMFGSNEGADGKQALPRLNAGLHTLAVDA 163
DB 145 SKAFPEMATLYGNTPNVYIATSPTECVLGGCOSS--EEVTAIRSIDPDGVVIGSP 202
QY 164 GNGQFQSIHDGREVFNADPQNTMFSIHMEYAGGNASQVNTIDRYLNODLAVIGE 223
DB 203 TMSQ-----DHLADNPVSHSNVYALHF--YSGTGGFLDRITTYANKGALFVTE 254
QY 224 FGHRTNGD-----VDPEATMSYSEQRGVGWLAMS 253
DB 255 WGTSDASGNGGPPYLPQSKEMIDFLNARKISWVWMS 289

RESULT 12

US-08-745-977-4
; Sequence 4, Application US/08745977
; Patent No. 5770406
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5770406 No. 5770406disk of No. 5770406th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745, 977
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4175.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-745-977-4

Query Match 6.1%; Score 100; DB 1; Length 429;

Best Local Similarity 21.9%; Pred. No. 0.11; Mismatches 119; Indels 88; Gaps 18;

QY 27 NHGHANYKQATTAIGSIANTGANTVRIYVLSDBGQWTKDIDHIVRN----- 72
DB 106 NHYRDWIN--PATVQSVHDVGLNTIRIP--GYWSYNAIVDTASBPFDAGNLQLPYLD 159

QY 73 -LISTLAEDNEHLVAPEVHDATG---YDSIASLN-----RAVDY-----WIMRSA 113
DB 160 AVYQKADLGITYIILDLHAPGGQQQDQDFTGQNPDPAGFYNTYDGRARKWLSWTKNRH 219
QY 114 LICKEDTV-IINIANEMFGSNEGADW--ADG-----YKQALPRLR-----NAGLN 155
DB 220 TNPATVSGKIEVLENPVSRHGGGRYPAPGQDPSKVQIYYFQALAAVDAEALNVPSN 279
QY 156 HTLVADAAKQGFQSIHDGREVFNADPQNTMFSIHMEYAG-----GNASQVR--- 206
DB 280 KRLHY-----QMSKMSQSDPRSNAAVKNPDMVEFDHNTYIGFALQNTGDDYSLMHS 333
QY 207 -INIDRYLNODLALVYGEGRHTNG-DVDEATMS-----YSEQRGVGWLAMSK 255
DB 334 CIDSRYVSGCDFE-ITGEWS--MTSGADWHDGNFTKFTTAQOQLYESPGMDGMIYTWK 390
QY 256 G--NGPEWEY 263
DB 391 TELNDPRWTY 400

RESULT 13

US-09-040-699A-4
; Sequence 4, Application US/09040699A
; Patent No. 6022723
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglucanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6022723 No. 6022723disk of No. 6022723th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040, 699A
; FILING DATE: 18-March-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valera A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4175.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-040-699A-4

Query Match 6.1%; Score 100; DB 3; Length 429;

Best Local Similarity 21.9%; Pred. No. 0.11; Mismatches 119; Indels 88; Gaps 18;

QY 27 NHGHANYKQATTAIGSIANTGANTVRIYVLSDBGQWTKDIDHIVRN----- 72

```

DB 106 NMYRMIN---PATVSHADVANTIRPI---GYMSNAYDTASEPFDQNLQPLYL 159
QY 73 -LISLAEDNHLVAVPEVDATG---YDSTIASLN-----RAVDY-----WLEKMSA 113
DB 160 AVYQKADIGIYIIDLHGAPEGQOQDAFTGQNPAGFYNTYDGRAEKMLSWMTNFIH 219
QY 114 LGKEDTV-IINANEMSGSWEGDAM--ADG-----YKQALPRLR-----NAGLN 155
DB 220 TNPASTVMIEVLNEPYSRHDGGRYPAPGQDPSPVQIYFGALKAVDAAALNVP SN 279
QY 156 HTLMDVAMGQFPOSINDYGREVNADPQRTMFSIMHYEAG-----GNASQVR--- 206
DB 280 KKLHV-----QFMSSKMGSGDPRSAVAKNDPMVGFDDHNYIGFALQNTGQYSLMISA 333
QY 207 -TVIDRYLMODIALVYGEGRHTNG-DVDEATIMS-----YSEBRGVGLAMSK 255
DB 334 CTDSPRVSGQDFA-TIGEMS--MTSGADWHDGNFTKFTTAQOOLYESPGMDGWLITWK 390
QY 256 G--NGPEWEX 263
DB 391 TELNDPRWTY 400

```

RESULT 14

US-09-066-075-2
Sequence 2, Application US/09066075

```

; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1744
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-066-075-2

```

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

```

QY 19 NPEYMGINHGHA-----W---YKQATTAIEGANTGATVRIYVSDSGQWTK--- 64
DB 9 NKILGRGINIGNMLEAPNEGDWGVYKDE--PEDIIKRGFSHVAKPI---KMSITHAV 61
QY 65 -----DDI--HTVKNLSLAEDNH-----LVAVEVDATGYDSIASINRA 103
DB 62 AEPPIKIMDFEKKVADVINGALKRGALVAINHTHYELANDNDEBEK-----ERF 111
QY 104 VDIWENKRSALIGKEDTVIINANEMSGSWEGDAMADGYQALPRLNAGLNTHTLWDA 163
DB 112 LALKQJADRYKQVETLFEELNEPFGNLTPEKWNELLEALVKVRSIDKKHTIIGTA 171
QY 164 GWGQFPQ---SIIDYGREVNADPQRTMFSIMY---EYAGSAGQVNTIDRYLND 216
DB 172 EWGGISALEKLSVPRW-----ENKSYVTHIYNPFEFTHOGALWVEGS--EKWIGRK 221
QY 217 LA-----LVIGEPGRHTNGVDYD-----ATIMSSEQ 246
DB 222 WGSPPDQKHLIEENFLEIEMSKKKRPIYIGEGF-AVRKADLESRIKWTFFVAREMEKR- 279
QY 247 VGLWMSKNGPEWEX 263
DB 280 ---RMSW---AYWEF 288

```

RESULT 15

US-08-518-615A-2
Sequence 2, Application US/08518615A

```

; Patent No. 5962258
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615A
; FILING DATE: August 23, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-518-615A-2

```

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:35:51 ; Search time 132.02 Seconds
(without alignments)
72.598 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGFYVSGTLLYDANGNP.....TIVNGPYGLRETSRLSTVF 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 692 | 42.5 | 383 | 1 MANA_STRLI | P51529 streptomyc |
| 2 | 601 | 36.9 | 1331 | 1 MANB_CALSA | P25533 cistridium |
| 3 | 202 | 12.4 | 409 | 1 GUN2_BACSA | P06565 bacillus sp |
| 4 | 201 | 12.3 | 400 | 1 GUN5_BACAG | P085465 bacillus ag |
| 5 | 200 | 12.3 | 488 | 1 GUN1_BACSA | P06566 bacillus sp |
| 6 | 184.5 | 11.3 | 505 | 1 GUNV_ERWCA | Q47096 erwilda car |
| 7 | 180.5 | 11.1 | 444 | 1 GUNV_ERWCA | Q59394 erwilda car |
| 8 | 180 | 11.0 | 448 | 1 GUN_CIOAB | P15704 cistridium |
| 9 | 177.5 | 10.9 | 466 | 1 GUN5_THRFU | Q01786 cistridium |
| 10 | 174.5 | 10.7 | 504 | 1 GUNV_ERWCA | Q59395 erwilda car |
| 11 | 173.5 | 10.6 | 499 | 1 GUN1_BACSU | P07983 bacillus su |
| 12 | 162.5 | 10.0 | 499 | 1 GUN2_BACSU | P10475 bacillus su |
| 13 | 161 | 9.9 | 459 | 1 GUN1_STRLI | P27035 streptomyc |
| 14 | 152 | 9.3 | 800 | 1 GUN_BACSA | P06564 bacillus sp |
| 15 | 151.5 | 9.3 | 499 | 1 GUN3_BACSU | P23549 bacillus su |
| 16 | 151.5 | 9.3 | 825 | 1 GUN3_BACSA | P19570 bacillus sp |
| 17 | 149.5 | 9.2 | 941 | 1 GUN_BACSA | P19424 bacillus sp |
| 18 | 147.5 | 8.7 | 429 | 1 GUN1_BACSA | P22541 butyridium |
| 19 | 142 | 8.7 | 562 | 1 GUN1_ACICE | P54583 acidotherm |
| 20 | 141 | 8.7 | 312 | 1 GUN4_RUMAL | Q07940 ruminococc |
| 21 | 131 | 8.0 | 814 | 1 GUN2_CIOAB | P10477 cistridium |
| 22 | 130 | 8.0 | 426 | 1 GUN_BACSA | P17974 burkholderi |
| 23 | 122.5 | 7.5 | 1039 | 1 GUNB_CALSA | P10474 cistridium |
| 24 | 122 | 7.5 | 438 | 1 EXG_CANAL | P29717 candida alb |
| 25 | 117 | 7.2 | 426 | 1 GUN2_ERWCH | P07103 erwilda chr |
| 26 | 115.5 | 7.1 | 747 | 1 GUN2_CERFI | P50400 cellulomona |
| 27 | 113.5 | 7.0 | 517 | 1 GUN4_CIOAB | P54937 cistridium |
| 28 | 111 | 6.8 | 475 | 1 GUNB_CIOAB | P17901 cistridium |
| 29 | 109.5 | 6.7 | 440 | 1 GUNB_CIOAB | P23660 ruminococc |
| 30 | 105 | 6.4 | 364 | 1 GUN4_RUMAL | P00533 cistridium |
| 31 | 104 | 6.4 | 566 | 1 GUN2_CIOAB | P27033 pseudomonas |
| 32 | 100.5 | 6.2 | 748 | 1 GUN2_PSEFI | P23776 pseudomonas |
| 33 | 99.5 | 6.1 | 448 | 1 EXG1_YEAST | P23776 pseudomonas |

| | | | | | |
|----|------|-----|-----|--------------|--------------------|
| 34 | 98 | 6.0 | 584 | 1 GUN2_CIOAB | P25472 cistridium |
| 35 | 97.5 | 6.0 | 375 | 1 GUN_FOBSF | P23044 robilliarca |
| 36 | 97.5 | 6.0 | 647 | 1 CN16_ECOLI | P08331 escherichia |
| 37 | 97 | 6.0 | 493 | 1 GUN4_XANCP | P19487 xanthomonas |
| 38 | 95 | 5.8 | 515 | 1 GUN2_CIOAB | P28623 cistridium |
| 39 | 94.5 | 5.8 | 645 | 1 LIPI_PHOLO | P40601 photorhabdu |
| 40 | 94.5 | 5.8 | 856 | 1 AAP1_YEAST | P37898 saccharomyc |
| 41 | 94 | 5.8 | 562 | 1 EXG2_YEAST | P52911 saccharomyc |
| 42 | 94 | 5.8 | 890 | 1 BCN5_CIOPE | P08696 cistridium |
| 43 | 93.5 | 5.7 | 406 | 1 GUN1_RUMAL | P16216 ruminococc |
| 44 | 93.5 | 5.7 | 510 | 1 YDID_SCHPO | O13692 schizosacch |
| 45 | 93.5 | 5.7 | 566 | 1 GUNB_PAEIA | P23550 penibacilli |

ALIGNMENTS

| RESULT | ID | MANA_STRLI | STANDARD | PRT | 383 AA. |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----------|-------------------------|---------|
| AC | P51529 | | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | | |
| DE | MANNAN ENDO-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-MANNANASE) (1,4-BETA-D-MANNAN MANNANOHYDROLASE). | | | | |
| GN | MANA. | | | | |
| OS | Streptomyces lividans. | | | | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | | | |
| CC | Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 36-42. | | | | |
| RC | STRAIN-66 / 1326; | | | | |
| RX | MEDLINE; 93207541. | | | | |
| RA | Arcand N., Kluepfel D., Paradis F.W., Morosol R., Shareck F.; | | | | |
| RT | "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence of the manA gene and characterization of the enzyme."; | | | | |
| RL | Biochem. J. 290:857-863(1993). | | | | |
| RN | [2] | | | | |
| RP | REVISIONS TO C-TERMINUS. | | | | |
| RC | STRAIN-66 / 1326; | | | | |
| RA | Shareck F.; | | | | |
| CC | Submitted (APR-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES CELSIUS AND A PH OF 6.8. | | | | |
| CC | -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS. | | | | |
| CC | -1- SUBUNIT: MONOMER. | | | | |
| CC | -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES). | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL; M92297; AAA26710.2; -. | | | | |
| DR | INTERPRO; IPR001547; -. | | | | |
| DR | PFAM; PF00150; cellulase.1. | | | | |
| DR | PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1. | | | | |
| FM | Hydrolase; Glycosylase; Signal. | | | | |
| FT | SIGNAL | 1 | 35 | | |
| FT | CHAIN | 36 | 383 | | |
| FT | DOMAIN | 336 | 340 | | |
| FT | SEQUENCE | 383 AA; | 39681 MW; | 5DBAB407C64E9AC3 CRC64; | |

Query Match 42.5%; Score 692; DB 1; Length 383;
Best Local Similarity 46.7%; Pred. No. 2,3e+46;

Matches 141: Conservative 51: Mismatches 104: Indels 6: Gaps 5:

OY 2 ANSGTYVSTIYDANGNFTVRGNGHNTKQATTAIESIANTGANTVIVISDGGQ 61
 DB 36 AAGGTHSVSGRVEENGSGSFVRGVNHNATWTPDR-TSISIDIAAKGANTVIVISGGR 94
 OY 62 WTKDDIHVFNRLISLAEDNHLVAPEVDATGY--DSIASLRADVNIEMSSAL-ICE 118
 DB 95 WTKTASAEVSTALIGOCKANKVYCVIEVDHTGYGQATSLQACDYVWGAASAMRAOE 154
 OY 119 DTVIINIANEWGSEGDAMADYQKQALPRLNAGLNLHTLVDAAGWQ-EPQSIHDYGR 177
 DB 155 DVVVNVIGNEPGNTNYAATDATKSAIGKLKAGLHMTVDAPVWQDMSGTRSNAA 214
 OY 178 EYFNADPQRTMFTSHMTEYAGNASQVRTNIDRYLNDLALVIEFGHRTNGVDAT 237
 DB 215 SYFASDPRTNVTFSIHMVG-VYDTLAEVVDYINAGVNGLPVVEGFGHSDGDDPDEA 273
 OY 238 IMSYEGRGVGLWMSKNGSGPEMEYLDLSNDWAGNLTAMGNTVNGPYGLRSTRST 297
 DB 274 IMATQSLGCVGLGMSGNGGCVETLDMVNGFDPNSLTSMGNTLLYSNGIATSKRTAT 333
 OY 298 VF 299
 DB 334 YV 335

RESULT 2
 MANB-CALSA STANDARD; PRT; 1331 AA.

AC P22533;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BETA-MANNANASE/ENDOGALACTANASE A PRECURSOR [INCLUDES: MANNAN ENDO-1,4-DE-MANNANOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-DE-MANNANASE); ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4) (CELLULASE)].
 GN MANA.
 OS Caldocellum saccharolyticum (Caldicellulosigrupor saccharolyticus).
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;
 RN Thermomonas group; Caldocellulosigrupor.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93119139.
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme.";
 RT Appl. Environ. Microbiol. 58:3864-3867 (1992).
 RN [2]
 RP SEQUENCE OF 1-346 FROM N.A.
 RX MEDLINE; 91247819.
 RA Luecht E., Jesmet N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'.";
 RT Appl. Environ. Microbiol. 57:694-700 (1991).
 CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGLUCANASE ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
 CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES CELSIUS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on the use of this data by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; L01257; AAA71887.1; -;
 DR EMBL; M35063; AAA72861.1; -;
 DR PIR; B43745; B43745.
 DR PIR; A48954; A48954.
 DR HSSP; Q06851; INBC.
 DR INTERPRO; IPR001547; -;
 DR INTERPRO; IPR001956; -;
 DR PRAM; PF00942; CBD_3; 2.
 DR PFAM; PF00150; cellulase, 1.
 DR PROSITE; PS00659; GLYCOSYL-HYDROL_5; 1.
 DR KEGG; K01001; Glycosylase; Cellulose degradation; signal;
 KW Multifunctional enzyme.
 FT SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT ACT_SITE 338 338
 FT CONFLICT 340 346
 SQ SEQUENCE 1331:AA; 146892 MM; FFBGCA51B8BDF0E0 CRC64;
 TTPPTPT -> ROHORO (IN REF. 2).
 T -> P (IN REF. 2).
 NDICLOPHILE (BY SIMILARITY).
 PROTON DONOR (BY SIMILARITY).
 CATALYTIC (ENDOGLUCANASE ACTIVITY).
 PRO/SER/THR-RICH (PT BOX).
 SUBSTRATE-BINDING (POTENTIAL).
 PRO/SER/THR-RICH (PT BOX).
 CATALYTIC (ENDOGLUCANASE ACTIVITY).
 PROTON DONOR (BY SIMILARITY).
 NDICLOPHILE (BY SIMILARITY).
 T -> P (IN REF. 2).
 TTPPTPT -> ROHORO (IN REF. 2).
 FFBGCA51B8BDF0E0 CRC64;

Query Match 36.9%; Score 601; DB 1; Length 1331;
 Best Local Similarity 44.5%; Pred. No. 1,4e-38;
 Matches 125: Conservative 42: Mismatches 108: Indels 6: Gaps 4:

OY 25 GINHGHWYDQATTAIESIANTGANTVIVISDGGQWTDHHTVNLISLAEDNHLVA 84
 DB 55 GTNHAHCWYDRDPTALRGIRSGWMSVRYVLSNGYMTWIRIPASEVANIISLSLGFKA 114
 OY 85 -VPEVDATGY--DSIASLRADVNIEMRSALIGKEDYIINIANEWGSEGDAMAD 140
 DB 115 IILEVHTTGYGEGGACSLAQAVETKESYLDGHEDEYIINIGEPYGNNTQWVN 174
 OY 141 GYKQALPRLNAGLNLHTLVDAAGWQ-EPQSIHDYGRVFNADPQRTMFTSHMTEYAG 199
 DB 175 DTKNAIRALRAGFKRTIMVDAPVWQDMSGTRMDNOSINEADPLNLVFSIHMVG-VY 233
 OY 200 GMSQVPTNIDRYLNDLALVIEFGHRTNGVDATIMSIEQGRVGLWMSKNGNP 259
 DB 234 NTASKVEEYKSFVQKLPVIEFGHRTNGVDATIMSIEQGRVGLWMSKNGNP 293
 OY 260 EWEYLDLSNDWAGNLTAMGNTVNGPYGLRSTRSTVPT 300
 DB 294 YVGILDMVNWMDRNPPTPGWQKTNALIGSSPTPTSTV 334

RESULT 3
 GUN2-BACS4 STANDARD; PRT; 409 AA.

AC P06565;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE).
 GN CELB.
 OS Bacillus sp. (strain N-4).
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE: 87056924.
RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alcalophilic
RL Bacillus sp. strain N-4 and their strong homology."
RC J. Bacteriol. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M14729; AAA22299.1; -.
DR PIR: B25156; B25156.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW Cellulose degradation; Hydrolase; Glycosidase.
FT ACCT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
FT ACCT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 409 AA; 43690 MW; 988860686B4DA3F CRC64;

Query Match 12.4%; Score 202; DB 1; Length 409;
Best Local Similarity 25.0%; Pred. No. 1.6e-08;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

QY 8 VSGTLLDANGNPFVARGI-NHGMAYKQATTAEGTANTGANTVRIYLSGGQWTKDD 66
DB 40 ISNGELVNRGEQVOLKGMSSHGLOWYGFVNYESMKWLRD 81
QY 67 -----IHT-----YANLSLADNHLVAVPEVDATGYDSIASLARA 103
DB 82 WGITVFRAMATISSGGYIEDPSKKEVKAENVAIDLIVITIDHIIISDNDPNITKKEA 141
QY 104 VDYWIEMRSALIGKEDVTYIINANEMFGS---WEGD--AMADGYKQATPRLNAGLNTL 158
DB 142 KDFPEEM-SELKGDYPNVITIELANEPNSDVTWNOIKPYAE---EVIPVIANNDPNNTI 197
QY 159 MVDAAGWQGFPOSIDYREVFNDPQRTNMFSSIMTEYAGGNASQVETNIDRVLNODLA 218
DB 198 IV---GTGTWSQDVH-HAADNQLTDP--NVMYAFHF--YAGTHGQNLRDQVYDALDQGA 249
QY 219 LVIGEFGRHRTNGD---VDENIT-MSYSEORGVGWLAWS 253
DB 250 IFVSEMGTSSEATGDGVFLDEAQAVIDFMDERNLSWANWS 289

RESULT 4
GUNS_BACAG STANDARD; PRT; 400 AA.
AC 085465;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENDOGLUCANASE 5A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (ALKALINE
DE CELLULOSE).
GN CELSA.
OS Bacillus agaradhaerens (Bacillus agaradhaerans).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 8721;
RA Bjornvad M.E.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE: 98153671.
RA Davies G.J., Dauter M., Brzozowski A.M., Bjornvad M.E.,
RA Andersen K.V., Schuelein M.;
RT "Structure of the Bacillus agaradhaerans family 5 endoglucanase at 1.6-
RT A and its cellobiose complex at 2.0-A resolution."
RL Biochemistry 37:1926-1932(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE: 98384136.
RA Davies G.J., MacKenzie L.F., Varrot A., Dauter M., Brzozowski A.M.,
RA Schuelein M., Withers S.G.;
RT "Snapshots along an enzymatic reaction coordinate: analysis of a
RT retaining beta-glycoside hydrolase."
RL Biochemistry 37:11707-11713(1998).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -----
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF067428; AAC19169.1; -.
DR PDB: 1A3H; 16-MAR-99.
DR PDB: 2A3H; 16-MAR-99.
DR PDB: 3A3H; 16-MAR-99.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACCT_SITE 165 165 PROTON DONOR.
FT ACCT_SITE 254 254 NUCLEOPHILE.
SQ SEQUENCE 400 AA; 44702 MW; 3F9C66FB9B36FF CRC64;

Query Match 12.3%; Score 201; DB 1; Length 400;
Best Local Similarity 26.6%; Pred. No. 1.8e-08;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLLDANGNPFVARGI-NHGMAY-----KQATTAEGTANTGANTVRI 54
DB 40 ISNGELVNRGEQVOLKGMSSHGLOWYGFVNYESMKWLRD-----DMGINVERA 89
QY 55 VL--SDGGQWTKDDI-HVIRNLISLADNHLVAVPEVDATGYDSIASLNRVADWIMER 111
DB 90 AMYTSSGGYIDDPYKKEVKAENVAIDLIVITIDHIIISDNDPNITKKEAFDEM- 148
QY 112 SALIGKEDVTYIINANEMFGSWEADADGYK---QAIPLRNAGLNTLNVDAAGWQ 167
DB 149 SELDYDYPNVITIELANEPNS--DVTWNOIKPYAEVITPIRNNDPNNITIV--GTGT 203
QY 168 FPOSIDYREVFNDPQRTNMFSSIMTEYAGGNASQVETNIDRVLNODLALVIGEFGR 227
DB 204 WSQDVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLRDQVYDALDQGAIFVSEMGTS 258
QY 228 HTNGD---VDENIT-MSYSEORGVGWLAWS 253
DB 259 AATGSGVFLDEAQAVIDFMDERNLSWANWS 289

RESULT 5
GUNI_BACSA4 STANDARD; PRT; 488 AA.
AC P06566;

```

```

DN 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
GN CELA.
OS Bacillus sp. (strain N-4).
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87056924.
RA Fukumori F., Sasahara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RT Bacillus sp. strain N-4 and their strong homology."
RL J. Bacteriol. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: M14781; AAA2301.1;
DR PIR: A25156; A25156.
DR INTERPRO: IPR001547;
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR KMW CELLULOSE degradation; Hydrolase; Glycosidase.
FT ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 488 AA; 54264 MW; 97248E453D28D3B5 CRC64;

```

Query Match 12.3%; Score 200; DB 1; Length 488;
 Best Local Similarity 24.6%; Pred. No. 2,8e-08;
 Matches 70; Conservative 49; Mismatches 103; Indels 62; Gaps 14;

```

OY 3 NSGFYSGTGLDANGNPNVNGI-NHGHWYKQDATTALIEGANTGATVRLVSDGG 61
DB 33 NGOLSTQNCQLVNEHGDVQLKGMSSHGQWYG-----GPNVDSIK 74
OY 62 WTKDD-----IHT-----YANLISLAEDNHLVAVPEVHATGYDSA 98
DB 75 WLRDDMGIVFRAMWTSSGGYIEDSVKRYEVAEVAIDGITYIIDMHLSDNDPI 134
OY 99 SLNRADVWIEMRSALIGKEDVIINANWFG---SWGDAMADY-KOALPRLNAGL 154
DB 135 YKEAEKEFDEEM-SALYGYPNVYIEIANEPNGHNVRW--DSHIKRYAEVIVIRANRP 191
OY 155 NHTLWADAGMGQFPOSIDYGRVFNADPQRTMFSIMHYEAGNAGASQVRNIDRYLN 214
DB 192 NNTVIVIGTATWS---ODVHEADNDQLD-DP--NVMYAFHF--YAGTHGOOLKQVYALIS 243
OY 215 QDLATIGEGHRTNGD---VDEATI-MSYSEQRGVGLWANS 253
DB 244 RGAATFVSEMGTSAAATGDGVFLDEAQAQWTDPEFDERNLSTANWS 287

```

RESULT 6
 GUNV_ERMCA STANDARD; PRT; 505 AA.

```

AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)
DE [CELLULOSE V].

```

```

GN CELV.
OS Eryinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SCRI193.
RX MEDLINE: 94067016.
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (celV) of Eryinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains."
RL Mol. Gen. Genet. 241:341-350(1993).
CC -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC CELSIUS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: X76000; CAA53592.1;
DR HSSP: O06851; INBC.
DR INTERPRO: IPR001547;
DR INTERPRO: IPR001956;
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR KMW CELLULOSE degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 505 ENDOGLUCANASE V.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;

```

Query Match 11.3%; Score 184.5; DB 1; Length 505;
 Best Local Similarity 23.1%; Pred. No. 4.6e-07;
 Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

```

OY 13 LYDANGNPFVNGI-NHGHWYKQDATTALIEGANTGATVRLVSDGGWTKDD----- 65
DB 48 LYDEQGRVQLRGISSHGLOMGD-----YANKDKMKWLRDDMGGINV 89
OY 67 -----IHYANLISLAEDNHLVAVPEVHATGYDSAISNRAVDYWI 108
DB 90 FRYAVMTADAGYISNPISLANRKEAVAAQSLGYITIIDMHLSDNDPIYKAAQKTFEA 149
OY 109 EMRSALIGKEDVIINANWFG---SWGD--AMADGYKQALPRLNAGLNTLWADAG 164
DB 150 EM-AGLIGSSPNVYIEIANEPNGVITWNGQIRPYA---LEVYDTRSDNDPDLIIV--G 202
OY 165 WGFPOSIDYGRVFNADPQRTMFSIMHYEAGNAGASQVRNIDRYLNQDLALVIGEF 224
DB 203 TGTWSDQIHD---AADNQPDPTMYALHF--YAGTHGOFLDRIDYQASRGAAIFVSEW 257
OY 225 GRRHNGD---VDEATIMSYSEQRGVGLWANSKNGKNGPEHYL-----DLSNDAGNLT 275
DB 258 GTSDASNGNGPPLPESQWTIDFLNNRGVSWVMSLTDSEASALAPGASKSGGWTEQNL 317
OY 276 TANGNTI 282

```

```

Db      318 STSGKFV 324
RESULT
7
GUNN_ERMCA          STANDARD:          PRT:    444 AA.
ID     GUNN_ERMCA
AC      Q59394;
Dt      01-NOV-1997 (Rel. 35, Created)
Df      01-NOV-1997 (Rel. 35, Last sequence update)
Dr      01-NOV-1997 (Rel. 35, Last annotation update)
DE      ENOGLUCANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE N)
        (CELLULOSE N).
GN      CELM.
OS      Erwiniacae carotovora.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC      Pectobacterium.
[1]
RN      SEQUENCE FROM N.A.
RC      STRAIN=ATROSEPTEICA FCBR C18;
RX      MEDLINE: 98299944.
RA      Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
        von Wettstein D.;
RT      "Transplanting two unique beta-glucanase catalytic activities into
        one multienzyme, which forms glucose.";
RL      Biotechnology 14:71-76(1996).
CC      -1- CATALYTIC ACTIVITY: ENDOPHROLYSTIS OF 1,4-BETA-D-GLUCOSIDIC
        LINKAGES IN CELLULOSE.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
        HYDROLASES).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC
DR      EMBL; L39788; AAC37033.1; -.
DR      INTERPRO; IPRO01547; -.
DR      INTERPRO; IPRO01566; -.
DR      Pfam; PF009442; CBD_3; 1.
DR      Pfam; PF00150; cellulase; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KM      Cellulose degradation; Hydrolase; glycosidase; signal.
FT      SIGNAL              1               31
FT      CHAIN              32             444
FT      ACCT SITE          168            168
FT      ACT SITE           256            256
FT      ACT SITE           256            256
SQ      SEQUENCE   444 AA: 48300 MW; FA7E4I79004CB843 CRC64;

Query Match      11.1%; Score 180.5; Db 1; Length 444;
Best Local Similarity 22.5%; Pred. No. 7.9e-07;
Matches 69; Conservative 48; Mismatches 123; Indels 67; Gaps 12;

13 LYDANGNPFWARGT-NHGAWTKDOATTAIESGANTGANTVAIVLSDGQWTKDD----- 66
Db      48 LVDEGGKRVQLRGVSHTGLWFGD-----TYMKDSKKMLRDPDWGINV 89
Oy      67 -----IHVRNTLSLAEDNHVLAVPEVDHATGDSTASLNRAVDWI 108
Db      90 FRVAYTTAADGTGISNPSLANKYKEVAQAAGSLGVIIITDHLLSDNDPRITYYAQAKTFFA 149
Oy      109 EMRSALLIGKEDTVIINIANWFQ--SWEGD--AMADGYKOALPRLNMAGLHTLVMDAAG 164
Db      150 EM-AELYGSSPVATIEINENPGVGVTWNGQIRPYA---LEVIDTIKSKDPDLIIV---G 202
Oy      165 WQGFPOSIIHDYGRVFVNADPQRNTWFSIMHYEYAGAGNSQVETINTIDRYLNDGLALVIGEF 224
Db      203 TETWSODIIDH---AADNQLPDPNTITALEHF--YAGTHGOFLDRIDYAQSRGAALFVSEW 257

```

[illegible]

DB 159 AEE-YGKSNVYIEICNEPNC---GTMNANDIKRYANYIIIPALIPDNNIIIVGTSTWS 214
 QY 167 QPPOSIMHYGAEVFNADPQR--NTMESIHMEYAGNASOVRTNIDRYLNODLALVIGEF 224
 DB 215 QDVADIAAD-----NPLRNSNMTYCHF--YAGTHQSILKDKINAMSGIAFTYEW 264
 QY 225 GHRRTNGD-----VDEA-TIMSYSOGRGVGLAWSKRGNGPMEYXL---DISNDWAGNNL 275
 DB 265 GTSDBSSGNGGPIYDSSQKWDVFMASKNISWTNNMALCDKSEASALKSGSSTTGWTDSDL 324
 QY 276 TAWGNTIYNGPYGLRETSRLS 296
 DB 325 TTSGLEFVKRSIGSGSNTTSOTS 345

RESULT 9
 GUNS_THERFU STANDARD; PRT; 466 AA.
 AC Q01786;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4)
 DE (CELLULOSE E-5) (CELLULOSE E5).
 GN CELE.
 OS Thermomonospora fusca.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptosporangiales; Nocardiopsaceae; Thermobifida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YX;
 RX MEDLINE; 91258320.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT "DNA sequences of three beta-1,4-endoglucanase genes from
 Thermomonospora fusca".
 RT J. Bacteriol. 173:3397-3407(1991).
 RN [2]
 RP REVISIONS.
 RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
 RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 137-142 AND 157-166.
 RC STRAIN-YX;
 RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
 RT "Activity studies of eight purified cellulases: specificity,
 synergism, and binding domain effects".
 RT Biotechnol. Bioeng. 42:1002-1013(1993).
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- PATHWAY: CELLULOSE DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: L01577; AAC09379.1; -
 DR PIR: C42360; C42360.
 DR HSSP: P07986; 1EXH.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001919; -
 DR PRAM: PRO0553; CBD.2; 1.
 DR PRAM: PRO0150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR Cellulose degradation; Hydrolase; Glycosylase; signal.
 KW SIGNAL
 FT CHAIN 1 36
 FT SIGNAL 37 466 ENDOGLUCANASE E-5.

FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 466 AA; 49800 KM; 1CF0ADBF2DEF82E CRC64;
 Query Match 10.9%; Score 177.5; DB 1; Length 466;
 Best Local Similarity 24.8%; Pred. No. 1,4e-06;
 Matches 78; Conservative 49; Mismatches 121; Indels 67; Gaps 16;
 QY 8 VSGTLLDANGNPFVAKGIN-HGHAWYKQAT-TALBGIA-NTGANTVRL---VLSDGQ 61
 DB 173 VCGTQLCDENGNPVLKRGMTSTHGIOWFDHCLTDSIDLALVDKAKADILRLSMYIQEDGYE 232
 QY 62 WT-----KDDIHTVANKLSLADNHLVAPREVHATGYSIASINRAVDYIEKRSALIG 117
 DB 233 TNPRTGFTDRNH---QLIDMTATANGLYIVDMHLLTPEDPHYNDPRATFTFAELTAQRASK 289
 QY 118 EDTVITINANEMFEGSWEGDAD--GY-KQALPRLNANGHNTLMDVDAAGWGFPOSIND 174
 DB 290 TN-VLYELANE---PQGVSWASIKSYAEVYIPYIRQDPDSVYIVGTRGMSIGVS-EG 343
 QY 175 YGREVFNADP--QRTMFSIHMEYAGNASOVRTNIDRYLNODLAL---VIGFGRHT 229
 DB 344 SGPAEIAANPNAASINIMYAEFFY-----AASHRDNYLNLREASELFPVFTVEFTET 397
 QY 230 NGD-----VDEATIMSYSOGRGVGLAWSK-----GNGPEYELDLS 267
 DB 398 TGDGANDFQMDR--YIDLMAERKIGMTKNYSIDPFSGAVFQPGTCASGP----- 447
 QY 268 NDWAGNNLTAWGNTI 282
 DB 448 --WSGSLKASGQWV 460

RESULT 10
 GUNW_ERMCA STANDARD; PRT; 504 AA.
 ID GUNW_ERMCA
 AC 059395;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ENDOGLUCANASE V1 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V1)
 DE (CELLULOSE V1).
 GN CELV1.
 OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SCC3193;
 RX MEDLINE; 95231512.
 RA Mae A., Helkinheimo R., Palva E.T.;
 RT "Structure and regulation of the Erwinia carotovora subsp. carotovora
 RT SCC3193 cellulase gene celv1 and the role of cellulase in
 RT phytopathogenicity".
 RT Mol. Gen. Genet. 247:17-26(1995).
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X79241; GAA55823.1; -
 DR HSSP: 006851; 1NRC.
 DR INTERPRO: IPR001547; -

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M16185; AAA2496.1; ALT_INIT.
DR PIR: A26874; A26874.
DR HSSP: 006851; INBC.
DR INTERPRO: IPR001547; -.
DR INTERPRO: IPR001956; -.
DR PFMF: PF00942; CBD_3; 1.
DR PFMF: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5_1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL
FT CHAIN 1 29
FT ACT_SITE 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55187 MW; 339D04EE95A63EE1 CRC64;

Query Match 10.6%; Score 173.5; DB 1; Length 499;
Best Local Similarity 20.8%; Pred. No. 3.2e-06;
Matches 69; Conservative 68; Mismatches 122; Indels 73; Gaps 15;

OY 3 NSGFVSGTTIYDANGNPFYKNGI-NHGAWYKQDATTALIGIANTGANTVRIYLSDGQ 61
DB 39 NGQISIKGTCLVNRDGRKAVQLKGSISSHGLQWYGD-----FVKKSDLK 80
OY 62 WTKDD-----IHT-----YRNLISLAEDNHLVAVPEVDATGYSIA 98
DB 81 WLWDMDGTFVFAAAAYTTADGGIDNPYKVKKEAVEAKELGIVIIDMLIDNGANQ 140
OY 99 SLNRADVWIEKRSALIKEDVTIININEMWEG--SFGED--ANADGYKQAIPLRNAGL 154
DB 141 NKEKKEKFFKEM-SLLYGNTPNVITVETLNERNGDVNMRKDKPAE---EYISVIRKNDP 196
OY 155 NHTLWDAAGWGQEPQSIHDYGREYFNADPQ--RNTMFSIHMEYTAGNAGNSQVARTNIDR 211
DB 197 DNIITV---GSGTWSQGVND-----AADQDLKANAVYALHF--YAGTHQGLSRKAN Y 245
OY 212 VLNDGLALVIGEFGRHNGD---VDEA-TIMTSYSEQRGVGLAWSKNGKPEWEYI-- 264
DB 246 ALSKAPAEFVEMWEGSDASGVFLDQSRMLNLDKSNISVWVNMVLSDKROESSALKP 305
OY 265 --DLSNDMAGNNLTFMAGNTIYNGPYGLRETSR 294
DB 306 GASKTGMPITDLTASGTFRNRTNGTKOSTK 337

RESULT 12
GPN2_BACSU STANDARD: PRT: 499 AA.
ID_GN2_BACSU P104/5.
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMC4SE) (CELLULOSE).
GN BGIC OR GMD OR EGLS.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PAP15;
RX MEDLINE; 87066783.
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moranelli F., Seligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RP [2]
RP SEQUENCE FROM N.A.
```

RC STRAIN-CK-2;
 RX MEDLINE: 95225656
 RA Landahl V., Aa K., Tronsmo A.;
 RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from *Bacillus subtilis* CK-2."
 RL Antonie Van Leeuwenhoek 66:327-332(1994);
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Rose M., Enliar K.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 30-45.
 RC STRAIN-CK-2;
 RX MEDLINE: 95225655.
 RA Aa K., Flengsrud R., Landahl V., Tronsmo A.;
 RT "Characterization of production and enzyme properties of an endo-beta-1,4-glucanase from *Bacillus subtilis* CK-2 isolated from compost soil."
 RL Antonie Van Leeuwenhoek 66:319-326(1994).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Z29076; CAA82317.1; -
 DR EMBL: X04689; CAA28392.1; -
 DR EMBL: X67044; CAA47429.1; -
 DR EMBL: Z73234; CAA97610.1; ALT_INIT.
 DR EMBL: Z99113; CAA13696.1; ALT_INIT.
 DR PIR: A26114; A26114.
 DR HSSP: Q06851; INBC.
 DR SUBTILIST: BG10437; BGIC.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001956; -
 DR PFAM: PF00942; CBD_3; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT CHAIN 1 29
 FT ACT_SITE 169 499
 FT ACT_SITE 257 257
 FT DOMAIN 350 499
 FT CONFLICT 283 283
 SQ SEQUENCE 499 AA; 55287 MW; 8F735FE711B3EAE2 CRC64;

Query Match 10.0%; Score 162.5; DB 1; Length 499;
 Best Local Similarity 21.0%; Pred. No. 2.2e-05;
 Matches 66; Conservative 75; Mismatches 136; Indels 37; Gaps 15;

QY 3 NSGFGVSGTLLYDANGNPFVAKGI-NHGHAWYKQDQAT-TAIEGIANTGANTV---RIVLS 57
 DB NGQSLSTGQTLVNRDGNKAVOLKSGISGLQVGYVKNKSLKRWLRDMGIVFRANAYTA 98
 QY 58 DGGQWTKDQI-HIVRNLSIAEDNHLVAVPEVHDATGYDSIASNRAVVDWIMERSALIG 116
 DB DGGYIDNPFVKNKYKAEVAKELGIYVILDMHILNDGNPNQNEKKEFFKEX-SSLYG 157
 QY 117 KEDVYIINIAEWFG--SMEDG--AMADGYKQAIPLRNAGLHFTLWDAAGMGQFQSI 172
 DB 158 NTPNVIYIINAEPPNDGVWKKDIKPYAE---EIVSVIRKNDPQIIVIT---GTSWQDV 211
 QY 173 HDYGRVFNADPO---RNTMFSIHMYEYAGNAGVQRTNIDRVINODLALVIGFGRHRT 229

DB 212 ND-----AADQLDANVATLHF--YAGTHGQFLRKANVATLSKGAPIVTEMGSDA 263
 QY 230 NGD-----VDEA-TINSYBORGCVGLAWSKNGEWEYTL-----DLSNDWAGNNULIANGN 280
 DB 264 SGNQGVFLDQSHKEMKLDKSKTISVWNWNLSDKQSSSALRFGASKTGWMLSDLSASGT 323
 QY 281 TIVNGPYGLRFSR 294
 DB 324 EVRENILGTKDSSTK 337

RESULT 13
 GUNA_STRLI STANDARD; PRT; 459 AA.
 ID GUNA_STRLI
 AC P27035;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE ENDOGLUCANASE CELA PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
 GN CELA.
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-43.
 RC STRAIN-66;
 RX MEDLINE: 92246492.
 RA DeBerge M., Lacaze P., Shareck F., Morosoli R., Kluepfel D.;
 RT "Purification and characterization of an endoglucanase from Streptomyces lividans 66 and DNA sequence of the gene."
 RT Appl. Environ. Microbiol. 58:815-820(1992).
 RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES IN CELLULOSE.
 CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL SITE FOR PROTEOLYSIS.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M82807; -; NOT_ANNOTATED_CDS.
 DR HSSP: P07986; LEXH.
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR001919; -
 DR PFAM: PF00553; CBD_2; 1.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00561; CBD_BACTERIAL; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 27
 FT CHAIN 28 459
 FT DOMAIN 136 135
 FT DOMAIN 147 147
 FT DOMAIN 148 357
 FT ACT_SITE 286 286
 FT ACT_SITE 378 378
 FT DISULFID 31 131
 SQ SEQUENCE 459 AA; 48663 MW; CFF47EC34E297A7 CRC64;

Query Match 9.9%; Score 161; DB 1; Length 459;
 Best Local Similarity 25.0%; Pred. No. 2.6e-05;
 Matches 81; Conservative 42; Mismatches 145; Indels 56; Gaps 16;

OY 3 NSGFVSGTTLTDANGNPFVARGIN-HGHAW-----YKDAATTAIGSIANTGANTVRIYLS 57
 DB 156 NGQHLVCGVHLCNOYDRPQLQNGSTHGIQWFGPCTGDSLDRLQDKMSDLIRVAMYQ 215
 OY 58 DGGQWTKDDIHT-VRNLSIAEDNHLVAVPEVDATGDSIASINRAVDYIEMRSALI 115
 DB 216 EDGYTDAGFTSRVNGSLVDAEDGMYAVIDFHLLTPDPYVNTDRATTF-----SSVA 271
 OY 116 GKED--TYIINIANFWEG-SWEG-DAMADGYKQAIPLRLNAGLNTLAVDAAGWQOFPOS 171
 DB 272 ARNDKKNVYIEIANEPNCSVTAVKSYAE---QVLPVIRADPDNAVIVTGRGMSL--G 326
 OY 172 IHDYRE---VFNADPQNTMFSTIMEFAGSNASQVPRINDRYLNQDLALYIGFGRH 228
 DB 337 VSDGANESEVANNPVNATINIMAFHF--YAASHKUDYRAAV-RPATRLPLFVSEFG--- 380
 OY 229 TNGVDENATIMSYSBQGVGLAW--SWKNGPEWEYLDL-----SNDMAGN 273
 DB 361 ----TVSATFANGVDSRSSVAMLDLDQKISYANTYSIDADGSAAFRCGTCGTDYSS 436
 OY 274 NLTANGNTVNGPYGLRSTRST 297
 DB 437 GVLTEGALVK-----SRIST 452
 RESULT 14
 GUN3_BACSI STANDARD; PRF: 800 AA.
 AC P06564:
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ENDOGUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (ALKALINE CELLULOSE).
 OS Bacillus sp. (strain 1139).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87085443.
 RA Fukumori F., Kudo T., Narahashi Y., Horikoshi K.;
 RT "Molecular cloning and nucleotide sequence of the alkaline cellulase
 gene from the alkalophilic Bacillus sp. strain 1139."
 RL J. Gen. Microbiol. 132:2329-2335(1986).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- MISCELLANEOUS: ALKALOPHILIC BACILLUS SP., STRAIN 1139, PRODUCES
 ONE ALKALINE CELLULOSE (PH OPTIMUM 9.0). THIS STRAIN IS NOT A TRUE
 CELLULOYTIC MICRO-ORGANISM BECAUSE THE ENZYME IS UNABLE TO
 HYDROLYSE NATIVE CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DB EMBL; M15743; AAA22305.1; -
 DB EMBL; D000066; BAA00045.1; -
 DR PIR; A29003; A29003.
 DR INTERPRO; IPR001547; -
 DR PFAM; PF00150; cellulase; 1.
 DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 30
 FT CHAIN 1 800
 FT ACT_SITE 190 190
 FT ACT_SITE 305 305
 FT NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 305 305

SQ SEQUENCE 800 AA; 88602 MW; 70CAAD786DAD55CF CRC64;
 Query Match 9.3%; Score 152; DB 1; Length 800;
 Best Local Similarity 22.8%; Pred. No. 0.00026;
 Matches 74; Conservative 53; Mismatches 125; Indels 72; Gaps 18;
 OY 12 TLTDANGNPFVARGIN-HGHAWYKQDAR-TAEGSIANT-GANTVRIYV--SNGQWTKDD 66
 DB 67 TLVDHGERIQLRGKSTGLQWFPFILNDNAKALANWESWIKILAVYVNGSASPE 126
 OY 67 IHTVNL--ISIAEDNHLVAVPEVH-DATGYDSIASINRAVDYIEMRSALIGKEDTVI 123
 DB 127 LKRSVYIGIDLAENDMVIYDVHVAHPGDRDYYAGAEFFDI-AALYPNPHIY 185
 OY 124 NIANFWEGSWEDA-----WADGYKQAIPLRLNAGL-NHTLVDAAGWQOFPO 170
 DB 186 ELANPESSNNNGAGIPNNEEGMNKVEYADPIVELRDSGNADNIIIVGSPNRSQRPD 245
 OY 171 SHIDYGREYVNDPQNTMFSTIMEFAGSNASQVPRINDRYLNQDL 217
 DB 246 LAAD-----NPIDHHTVYVFTGSHAATESYPPETPNSERNVMSNTRALENGY 299
 OY 218 ALVIGFGRHRTNGD---VDEATI-MSYSBQGVGLAWS-----WKG 256
 DB 300 AVFATEWGTSCANGDGGPFPEADAVWIEFLNENNISMANSILTNKNEVSGATTPELGKS 359
 OY 257 N-----GPE-W--EYLDLSNDW 270
 DB 360 NATSLDPGPDQVWPEELSLSGEY 383
 RESULT 15
 GUN3_BACSU STANDARD; PRF: 499 AA.
 AC P23549:
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE ENDOGUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CARBOXYMETHYL-CELLULOSE) (CMCSE) (CELLULOSE).
 GN BGIC.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BSE616.
 RX MEDLINE; 91299280.
 RA Park S.H., Kim B.K., Park M.Y.;
 RT "Characterization and structure of the cellulase gene of Bacillus
 subtilis BSE616."
 RL Agric. Biol. Chem. 55:441-448(1991).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DB EMBL; D01057; BAA00859.1; -
 DR PIR; JN0111; JN0111.
 DR HSSP; Q06851; INBC.
 DR INTERPRO; IPR001547; -
 DR INTERPRO; IPR001956; -
 DR PFAM; PF00942; CBD_3; 1.
 DR PFAM; PF00150; cellulase; 1.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:06:42 ; Search time 248.55 Seconds
(without alignments)
76.597 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630

Sequence: 1 NANSGFYVSGTITLYDANGP.....TIVNGPYGLRSLSTVFF 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR65:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 1169 | 71.7 | 516 | 2 JE0134 | mannan endo-1,4-be |
| 2 | 710 | 43.6 | 384 | 2 T37232 | secreted beta-mann |
| 3 | 692 | 42.5 | 363 | 2 S30386 | mannan endo-1,4-be |
| 4 | 601 | 36.9 | 1331 | 2 A48954 | mannan endo-1,4-be |
| 5 | 202 | 12.4 | 409 | 2 B25156 | cellulase (EC 3.2. |
| 6 | 200 | 12.3 | 488 | 2 A25156 | cellulase (EC 3.2. |
| 7 | 197 | 12.1 | 557 | 2 JC5487 | cellulase (EC 3.2. |
| 8 | 184.5 | 11.3 | 505 | 2 S39662 | endoglucanase - Er |
| 9 | 180 | 11.0 | 448 | 2 A27631 | cellulase (EC 3.2. |
| 10 | 176.5 | 10.8 | 466 | 2 CA2360 | cellulase (EC 3.2. |
| 11 | 174.5 | 10.7 | 499 | 2 A27198 | cellulase (EC 3.2. |
| 12 | 174.5 | 10.7 | 504 | 2 S34744 | cellulase (EC 3.2. |
| 13 | 173.5 | 10.6 | 508 | 2 A26874 | cellulase (EC 3.2. |
| 14 | 162.5 | 10.0 | 486 | 2 I40548 | bifunctional cellu |
| 15 | 162.5 | 10.0 | 508 | 2 G69593 | cellulase (EC 3.2. |
| 16 | 157 | 9.6 | 783 | 2 JC5467 | cellulase (EC 3.2. |
| 17 | 152 | 9.3 | 800 | 2 A29003 | cellulase (EC 3.2. |
| 18 | 152 | 9.3 | 822 | 2 JT0611 | cellulase (EC 3.2. |
| 19 | 151.5 | 9.3 | 499 | 2 JN0111 | cellulase (EC 3.2. |
| 20 | 151.5 | 9.3 | 825 | 2 JS0174 | cellulase (EC 3.2. |
| 21 | 149.5 | 9.2 | 357 | 2 PC4404 | cellulase (EC 3.2. |
| 22 | 149.5 | 9.2 | 941 | 2 S29043 | cellulase (EC 3.2. |
| 23 | 147.5 | 9.0 | 429 | 2 S29044 | endoglucanase A pr |
| 24 | 136.5 | 8.4 | 570 | 2 S56132 | cellulase (EC 3.2. |
| 25 | 131 | 8.0 | 814 | 1 C6CLEM | cellulase (EC 3.2. |
| 26 | 130 | 8.0 | 426 | 2 A42649 | cellulase (EC 3.2. |
| 27 | 127 | 7.8 | 32 | 2 PC4278 | guar gum-degrading |
| 28 | 122.5 | 7.5 | 915 | 2 A43802 | cellulase (EC 3.2. |
| 29 | 122.5 | 7.5 | 1039 | 2 S02711 | cellulase (EC 3.2. |

| | | | | | |
|----|-------|-----|-----|----------|--------------------|
| 30 | 122 | 7.5 | 438 | 2 A47702 | glucan 1,3-beta-gl |
| 31 | 117.5 | 7.2 | 428 | 2 S03767 | cellulase (EC 3.2. |
| 32 | 116 | 7.1 | 611 | 2 JC7177 | endoglucanase V (E |
| 33 | 115.5 | 7.1 | 747 | 2 B47093 | cellulase (EC 3.2. |
| 34 | 114 | 7.0 | 419 | 2 S72325 | glucan 1,3-beta-gl |
| 35 | 113.5 | 7.0 | 517 | 2 I40798 | cellulase (EC 3.2. |
| 36 | 111.5 | 6.8 | 441 | 2 A44815 | cellulase (EC 3.2. |
| 37 | 111 | 6.8 | 475 | 1 C6CLCA | cellulase (EC 3.2. |
| 38 | 110 | 6.7 | 26 | 2 PC4285 | guar gum-degrading |
| 39 | 109.5 | 6.7 | 754 | 2 T14877 | hypothetical prote |
| 40 | 107 | 6.6 | 430 | 2 S55325 | endo-beta-1,6-gluc |
| 41 | 105 | 6.4 | 364 | 2 S12017 | endoglucanase A - |
| 42 | 105 | 6.4 | 456 | 2 T40276 | probable gaal faml |
| 43 | 104 | 6.4 | 566 | 2 A40589 | cellulase (EC 3.2. |
| 44 | 101.5 | 6.2 | 317 | 2 B72216 | endoglucanase - Th |
| 45 | 100.5 | 6.2 | 748 | 2 S19652 | cellodextrinase C |

ALIGNMENTS

RESULT 1
JE0134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N:Alternate names: endo-1,4-beta-mannanase
C:Species: Bacillus circulans
C:Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C:Accession: JE0134
R:Yoshida, S.; Sako, Y.; Uchida, A.
BioSci. Biotechnol. Biochem. 62, 514-520, 1998
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: JE0134; MUID:98233274
A:Accession: JE0134
A:Molecule type: mRNA
A:Residues: 1-516 <YOS>
A:Cross-references: DDBJ:AB007123
A>Note: the authors translated the codon CAA for residue 259 as Asn, CAA for residue
A>Note: the translation of the nucleotide sequence 1294-1365 is not given in this seq
C:Keywords: glycoprotein; glycosidase; hydrolase

| | | | | |
|-----------------------------------------------------------------------------|------------------|------------------------------------------------------------|-------------------------------------|-------------|
| Query Match | 71.7% | Score 1169: | DB 2, | Length 516: |
| Best Local Similarity | 69.2% | Pred. No. 1e-83: | | |
| Matches 207: | Conservative 48: | Mismatches 44: | Indels 0: | Gaps 0: |
| QY | 1 | NANSGFYVSGTITLYDANGNP | FYMRGINHGMYKQDATTALIEGIANTGVRIYLSDG | 60 |
| DB | 33 | HAASGFYVSGTITLYDANGNP | FYMRGINHGMYKQDATTALIEGIANTGVRIYLSDG | 60 |
| QY | 61 | QNTKDIHTVKNLISLAEDNHLVAVPEVHDATGYDSIASLNRAVDYWMRSALIGKEDT | | 120 |
| DB | 93 | KWTLDDVNTVNNITLTCRQNKILAVLEVHDATGSSLSLDLNAVYWGIRKALIGKEDR | | 152 |
| QY | 121 | VIIINIANEWFSGWEDDANADGKOAIPRLRNAGLHTLTMVDAAGGQRPDSYKNGTEVL | | 180 |
| DB | 153 | VIIINIANEWFSGWEDDANADGKOAIPRLRNAGLHTLTMVDAAGGQRPDSYKNGTEVL | | 212 |
| QY | 181 | NADPONTWFSIHTMEYAGNAGQVNTINDRYLNODLALVIEFGHRTNGVDYDATTIS | | 240 |
| DB | 213 | NADPONTWFSIHTMEYAGNAGQVNTINDRYLNODLALVIEFGHRTNGVDYDATTIS | | 272 |
| QY | 241 | YSEQKGVGLANSMKNGGPEWEYDLSDMAGNNLTAMGNTIVNGPYGLRSLSTVFF | | 299 |
| DB | 273 | YSEKGVGLANSMKNGGPEWEYDLSDMAGNNLTAMGNTIVNGPYGLRSLSTVFF | | 331 |
| RESULT 2 | | | | |
| T37232 | | secreted beta-mannosidase - Streptomyces coelicolor | | |
| C:Species: Streptomyces coelicolor | | | | |
| C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999 | | | | |
| C:Accession: T37232 | | | | |
| R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. | | | | |

submitted to the EMBL Data Library, September 1998

A:Reference number: 221615

A:Accession: T37232

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-364

A:Cross-references: EMBL:AL031514; PIDN:CAA20610.1; GSPDB:GNO0070; SCOEDB:SC2H4.16

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: manA2; SCOEDB:SC2H4.16

Query Match

Best Local Similarity 46.7%; Score 710; DB 2; Length 384;

Matches 141; Conservative 53; Mismatches 102; Indels 6; Gaps 4;

```

QY 2  AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIVLSDGQ 61
DB 37  AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIVLSDGQ 61
QY 62  WTAKDIIHTVRLNLSLAEDNHLVAVPEVHDATGY--DSIASLNRADVYIEMRSALIGKE 118
DB 96  WTAKDIIHTVRLNLSLAEDNHLVAVPEVHDATGY--DSIASLNRADVYIEMRSALIGKE 118
QY 119  DTVIINIANEMEGSWEGDAMADYKQALPRLNAGLNHTLMDVADAGWQ-FPOSIDHYGR 177
DB 156  DTVIINIANEMEGSWEGDAMADYKQALPRLNAGLNHTLMDVADAGWQ-FPOSIDHYGR 177
QY 178  EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGVDENAT 237
DB 216  EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGVDENAT 237
QY 238  IMSYBQRCGVGLWMSWKGNGPEWEYLDLSNDWAGNNLTAMGNTIVNGPYGLRETSRLST 297
DB 275  IMATQSLGVGYLGWSWGSNGGVEYLDLVNGDFDPSLISWGNRIYSGNGIAATSRTAT 334
QY 298  VF 299
DB 335  VF 336

```

RESULT 3

manan endo-1,4-beta-mannosidase (EC 3.2.1.78) precursor - Streptomyces lividans

N:Alternate names: beta-mannanase

C:Species: Streptomyces lividans

C:Date: 02-Dec-1993 #sequence_revision 13-Mar-1997 #text_change 22-Oct-1999

R:Accession: S30386; S27699

R:Arcand, N.; Kuepfel, D.; Paradis, F.W.; Morosoli, R.; Shareck, F.

Biochem. J. 290, 857-863, 1993

A:Title: beta-Mannanase of Streptomyces lividans 66: cloning and DNA sequence of the man

A:Reference number: S30386; MUID:93207541

A:Accession: S30386

A:Molecule type: DNA

A:Residues: 1-363 <ARC>

A:Cross-references: EMBL:M92297; NID:g153193; PIDN:AAA26710.1; PID:g404076

A:Experimental source: strain 1326

C:Genetics:

A:Gene: manA

C:Function:

A:Description: catalyses degradation of beta-mannans into short-chain oligo-mannosides;

A:Pathway: mannose metabolism

A:Note: Industrially used for bioleaching of wood pulps

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-36/Domains: signal sequence #status predicted <SIG>

F:37-363/Product: mannan endo-1,4-beta-mannosidase #status predicted <MAT>

Query Match

Best Local Similarity 42.5%; Score 692; DB 2; Length 363;

Matches 141; Conservative 51; Mismatches 104; Indels 6; Gaps 5;

QY 2 AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIVLSDGQ 61

```

DB 36  AAGGIVSGTTLTDANGNPFVRNGINHGAMTKDQATTAIEGANTGANTVRIVLSDGQ 94
QY 62  WTAKDIIHTVRLNLSLAEDNHLVAVPEVHDATGY--DSIASLNRADVYIEMRSAL-IGKE 118
DB 96  WTAKDIIHTVRLNLSLAEDNHLVAVPEVHDATGY--DSIASLNRADVYIEMRSAL-IGKE 118
QY 119  DTVIINIANEMEGSWEGDAMADYKQALPRLNAGLNHTLMDVADAGWQ-FPOSIDHYGR 177
DB 156  DTVIINIANEMEGSWEGDAMADYKQALPRLNAGLNHTLMDVADAGWQ-FPOSIDHYGR 177
QY 178  EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGVDENAT 237
DB 216  EYFNADPQNTMFSIMHEYAGNAGSOYRTNIDRYLNODLALVIGFGRHRTNGVDENAT 237
QY 238  IMSYBQRCGVGLWMSWKGNGPEWEYLDLSNDWAGNNLTAMGNTIVNGPYGLRETSRLST 297
DB 274  IMATQSLGVGYLGWSWGSNGGVEYLDLVNGDFDPSLISWGNRIYSGNGIAATSRTAT 333
QY 298  VF 299
DB 334  VF 335

```

RESULT 4

manan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum

N:Alternate names: beta-mannanase

C:Species: Caldocellum saccharolyticum

C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999

R:Accession: A48954; B43745

R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.

Appl. Environ. Microbiol. 58, 3864-3867, 1992

A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multimer

A:Reference number: A48954; MUID:93119139

A:Accession: A48954

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-1331 <GIB>

A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291

A:Note: Sequence extracted from NCBI backbone (NCBI:121576, NCBI:121577)

R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.

Appl. Environ. Microbiol. 57, 694-700, 1991

A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene co-

A:Reference number: A43745; MUID:91247819

A:Accession: B43745

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-337, 'PROHORO' <LUE>

A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294

A:Note: The authors translated the codon CAC for residue 262 as Glu

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.9%; Score 601; DB 2; Length 1331;

Best Local Similarity 44.5%; Pred. No. 7,36-39;

Matches 125; Conservative 42; Mismatches 108; Indels 6; Gaps 4;

```

QY 25  GINHGAMTKDQATTAIEGANTGANTVRIVLSDGQNTKDDIHTVRLNLSLAEDNHLVA 84
DB 55  GINHGAMTKDQATTAIEGANTGANTVRIVLSDGQNTKDDIHTVRLNLSLAEDNHLVA 84
QY 85  -VPEVHDATGY--DSIASLNRADVYIEMRSALIGKEDYIINIANEMEGSWEGDAMAD 140
DB 115  ILEVHDATGY--DSIASLNRADVYIEMRSALIGKEDYIINIANEMEGSWEGDAMAD 140
QY 141  GYKQALPRLNAGLNHTLMDVADAGWQ-FPOSIDHYGRVFNADPQNTMFSIMHEYAG 199
DB 175  DTKNIAKLRLDAGFHTTMDVADAPRWQDWSMTMDNAGSIEADPLRLVYSIMHTG-VY 233
QY 200  GNAGSVRTNIDRYLNODLALVIGFGRHRTNGVDENATIMSYSBQRCGVGLWMSWKGNGP 259

```

D_b 234 NTASRVEYIISFPDKGIPLYIGEFGRHDHDDDEALIRYAQYRIAGLEFSMSWGSS 239

Q_Y 260 EMEYLDLSNDWAGNNLTAWGNTLYNGPYGLBETSRLSVFT 300

D_b 294 YVGYLDMVNNMDPRNNPTFGWGMYKTNALGISSTPTPTSYVT 334

| RESULT | 5 |
|--------|---|
|--------|---|

cellulase (EC 3.2.1.4) 2 - *Bacillus* sp.
N:Alternate names: endo-1,4-beta-glucanase
C:Species: *Bacillus* sp.
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: B25156
R:Fukunori, F.; Sasahara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168, 479-485, 1986
A:Title: Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain
A:Reference number: A91825; M01D:87056924
A:Accession: B25156
A:Molecule type: DNA
A:Residues: 1-409 <FUK>
A:Cross-references: GB:M14729; NID:g142655; PIDN:AAA22299.1; PID:g142656
A:Experimental source: strain N-4, plasmid pNK2
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

| | | | | |
|-------------|--------|------------|-------|-------------|
| Query Match | 12.48; | Score 202; | DB 2; | Length 409; |
|-------------|--------|------------|-------|-------------|

Best Local Similarity 25.0%; Pred. No. 2.1e-08;
Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

```

Oy      8 VSGTLLYDANGNPFVWIRGI-NEGHWAKKQDATTAEIGIANTGANTYRIYLSDGQWTKDD 66
      :| | : | | ::| | | | : | : | |
Db     40 ISNGELVNRGEPVQJLKGMSISGLQWIG-----QFVNTSEMKRLRDD 81

```

```

Oy 67 -----IHT-----VRLISLAEADNHLVAPPEVHDATGYDSIASLNR 103
      ::|               | : | : | : | : | : |
Db 82 WGIIVFRRAAMYTSSGCIEDPSYKKEKYEAVEAIDLGIYIIDWHILSDNDPNITYKEA 141

```

[illegible]

```

QY      219 LVIGFEGHRRHNGD---VD EATI-MSYSEQRGVGNLAWs 253
          :::::
Db      250 IFVSEWGTSEATG DGVFLDEA QVWIDFMDERNLsWAWNS 289

```

RESULT 6

cellulase (EC 3.2.1.4) 1 - *Bacillus* sp.
N/Alternate names: endo-1,4-beta-glucanase
C/Species: *Bacillus* sp.
C/Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C/Accession: A25156
R/Fukunori, F.; Sashihara, N.; Kudo, T.; Horikoshi, K.
J. Bacteriol. 168: 479-485, 1986
A/Title: Nucleotide sequences of two cellulase genes from alkaliphilic *Bacillus* sp. strain
A/Reference number: A91825; M01D:87056924
A/Accession: A25156
A/Molecule type: DNA
A/Residues: 1-488 <FUK>
A/Cross-references: GB:M14781; GB:X53449; NTD:g142659; P1DN:AAA22301.1; PTD:g142660
A/Experimental source: strain N-4, plasmid pNK1
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-

C; Keywords: glycosidase; hydrolase; polysaccharide degradation

| | | | | |
|--------------------------|-------|--------------------|------------|-------------|
| Query Match .. | 12.3% | Score 200; | DB 2; | Length 488; |
| Best Local Similarity | 24.6% | Pred. No. 3.8e-08; | | |
| Matches 70; Conservative | 49; | Mismatches 103; | Indels 62; | Gaps 14; |

```

D5 33 NGQULSQNGQLNENSGDFVQLNGMSSBGLQMG-----QFVWIDSK 74
QY 62 WTKOD-----IHT-----VRNLSLAEDNHLVAPEVHDATGYSIA 98
   | : || | : || | : || | : || | : || | : || | : || | : ||
D5 75 WLRODMGIVTIRAAITSSGTYEDBPVKEKYEKVEALDGLITVILDHILSDPNI 134
QY 99 SLNRAVDWIMERSALKIKEDVTIINIANEWFG--SNEGADMGY-KQAPRLNAGL 154
   | : || | : || | : || | : || | : || | : || | : || | : ||
D5 135 YKEAKEFEFDEI-SALYGDYPNVITYEIANEPNGHNVR--DSHIKPYAEVILPIRANDP 191
QY 135 NHTLMVDAAGNGQFPQSIHDYGRYEVFNADPQARNTFSLHXYEYAGNNSQVRTINDRYLN 214
   | : || | : || | : || | : || | : || | : || | : || | : ||
D5 192 NNIVYVGTATWS--QDVHEADNQLD-DP--NWTATHF--YAGTHQOQLNRNOVDYALS 243
QY 215 ODIALVIGEFGRHTNGD---VDPATI-MEYSRGQGYGLAWS 253
   | : || | : || | : || | : || | : || | : || | : || | : ||
D5 244 RGAALFVSEKGTSAATGDGVFLDRAQWIDIPKBERNLSTANWS 287

```

RESULT 7

uc35487
 cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
 N:Alternate names: endo-1,4-beta-glucanase
 C:Species: Clostridium cellulovorans
 C:Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 15-Oct-1999
 C:Accession: J05487, PC4333
 R:Shewfelt, S.A.; Ichl-Ishl, A.; Park, J.S.; Liu, C.; Malburg, L.M.; Dol, R.H.
 Gene 182, 163-167, 1996
 A>Title: Characterization of engF, a gene for a non-cellulosomal Clostridium cellulov
 A:Reference number: J05487; MUID:97135706
 A:Accession: J05487
 A:Molecule type: DNA
 A:Residues: 1-557 <SHE1>
 A:Cross-references: GB:J37056; NID:g1778708; PIDN:AAB0891.1; PID:g1778709
 A:Experimental source: strain ATCC 35296
 A:Accession: PC4333
 A:Molecule type: protein
 A:Residues: 30-37 <SHE2>
 C:Comment: This enzyme plays a secondary role in cellulose degradation.
 C:Genetics:
 A:Gene: engF
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F.I.29/Domain: signal sequence #status predicted <SIG>
 /30-557/Product: cellulase #status predicted <CAT>

| Query Match | Score | DB 2 | Length | 557; |
|-------------|-------|------|--------|------|
|-------------|-------|------|--------|------|

Best Local Similarity -26.78; Pred. No. 7.7e-08;
Matches 88; Conservative 33; Mismatches 132; Indels 76; Gaps 17;

```

Oy      12 TLVANGNPFYMRGIN-HGHAMTKDQATAIEGIAN TG-----NTVRVYL--SD 58
        || :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      62 TLCKDKGNPIQLRGSMTHGLQW-----PGVVNNNAFALSNDMSNVIRLAMYAE 11

```

```

0Y 59 GGGTCKDDI-HTVRNLSLAEDNHLVAPEVHDATGYDSISLNRAVDYWEKMSALIGK 11
    | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 114 GGYATNPSCVTYINGINTALANDMTYLYDWHMNPDPNNSVYSGAQSFENDISTLYPN 17

```

```
Oy      118 EDYVIANIAEWFGSGWEG---DA--WADGYKQALP---RLRNAGLNTHTLMVDAAWGQGF 16
       : : | | | | | | | | | | | | | | | | | |
Db      174 NKNIIELCNEPNGENGGVTNDATGMAOVKSYPATIVOLLNRKGNGENMLITVGPEWSORP 23
```

QY 170 QSHDYGREVNADPQNTMFSIHMEYAGCN-ASQVRTNDR-----VLNODLALVI 22

Db 234 DLAAD-----NPINDSNMTSVH--YSGNPISITVDINRDNMSNVRYALNNGHAIFYA 285
 QY 222 GEGHRTNGVDVDEATINSYSEQRGVGLWLSWKGNGPEWEYLDSND----- 269
 Db 286 TEMGTSIATG-----TTPPYL-ARKADAWLDF-LNGNINISMCNFSISNDEKRAAALNSITS 338
 QY 270 -----WAGNNLWAGNTI---VNGPY 287
 Db 339 LDPSGDKLMADNELITTSQYVRAKIKGAY 367

RESULT 8

339962
 endoglucanase - *Erwinia carotovora*

C:Species: *Erwinia carotovora*
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
 C:Accession: S39962

R:Cooper, V.J.C.; Salmond, G.P.C.
 M:1. Gen. Genet. 241, 341-350, 1993

A:Title: Molecular analysis of the major cellulase (CelV) of *Erwinia carotovora*: evidenc
 A:Reference number: S39962; MUID:94067016

A:Accession: S39962
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-505 <COO>

A:Cross-references: EMBL:X76000; NID:9434941; PIDN:CAA53592.1; PID:9434942

Query Match 11.3%; Score 184.5; DB 2; Length 505;
 Best Local Similarity 23.1%; Pred. No. 6.4e-07;
 Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

QY 13 LYDANGNFVNRGI-NHGHAWYKQATTAIGSINTGANTYRIYLSDDGQTKPD----- 66
 Db 48 LYDEGKRVQLRGSSRGQLWFGD-----YVKKDSMKVLRODMGINTV 89
 QY 67 -----IHTVRNLISLAEDNHLVAPEVHDAGYDSIASLNRADVWI 108
 Db 90 FRVAMTAAADYISNPLANKYKEVAAQSIGYIIITDHLISDNDNITKAQAKTFPA 149
 QY 109 EMRSALICKEDTYIINANWFG--SWEGD--AWADYKQAIPLRLNAGLHNTLWVDAAG 164
 Db 150 EM-AGLYSSPNVYIYEINENPGGVWNGQIRPYA---LEVTDTRSKDPNLTIV---G 202
 QY 165 WGOPOSTHDGREFVNDPQRTMESIHMYEYAGNAGNSQVRTINDRLNDLALVIGEF 224
 Db 203 TGTWSQDTHD--AADNOLPDPNTYALHF--YAGTHQQLRDRIDYVAGSAGALFVSEW 257
 QY 225 GHRHTNGD-----VDEATINSYSEQRGVGLWLSWKGNGPEWEYL---DLSNDWAGNLT 275
 Db 258 GTSASGNGGFLPESQWIDFNNRGYSWVNSLTDKSEASALAPGASKSGGWTGQNL 317
 QY 276 TANGNTI 282
 Db 318 STSGKTV 324

RESULT 9

A27631
 cellulase (EC 3.2.1.4) precursor - *Clostridium acetobutylicum*

N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Clostridium acetobutylicum*

C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
 C:Accession: A27631

R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
 Appl. Environ. Microbiol. 54, 1289-1292, 1988

A:Title: Structure of an endo-beta-1,4-glucanase gene from *Clostridium acetobutylicum* P2

A:Reference number: A27631; MUID:88268074
 A:Accession: A27631

A:Molecule type: DNA
 A:Residues: 1-448 <ZAP>

A:Cross-references: EMBL:M31311; NID:9144789; PIDN:AAA23230.1; PID:9144790

A:Note: the authors translated the codon GAG for residue 116 as Gly, GAA for residue 263
 9, 75

C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 11.0%; Score 180; DB 2; Length 448;
 Best Local Similarity 22.1%; Pred. No. 1.2e-06;
 Matches 71; Conservative 59; Mismatches 135; Indels 56; Gaps 14;

QY 8 VSGTTLVDANGNPFVNRGI-NHGHAWY-----KQATTAIGSINTGANTVRI 54
 Db 49 VVSGQLCDNSGNKPIQLKGMSSHGLWYVNFVNDKMKFLDK-----MGVNVIRA 98
 QY 55 VL--SDGGQWT--KDIHTVRLNLSLAEDNHLVAPEVHDATGYDSIASLNRADVWTEM 110
 Db 99 AMYINTEGGYISNPSOKEKTKIYQDAIDIMVYIIDMHLISNPNPTYEKQAKSFEQEM 158
 QY 111 RSALIGKEDTYIINANWFGNSWEGDAMADYK---QAIPLRLNAGLHNTLWVDAAGW 166
 Db 159 AEF-YGKISNVYIEICNEPNG---GYNWANDIKPYANTYIIPALRAIDPNNTITVGSTWS 214
 QY 167 QPQSHIDYGRVFNADPOR--NTMESIHMYEYAGNAGNSQVRTINDRLNDLALVIGEF 224
 Db 215 QDVIDIAD-----NPLRYSNIMYTCHE--YAGTHQSLRDKINYMKGIAIFVTEW 264
 QY 225 GHRHTNGD-----VDEA-TINSYSEQRGVGLWLSWKGNGPEWEYL---DLSNDWAGNLT 275
 Db 265 GTSASGNGGFLPESQWIDFNNRGYSWVNSLTDKSEASALAPGASKSGGWTGQNL 324
 QY 276 TANGNTIYNGPYGLRSTRLS 296
 Db 325 TTSGLFYKKSIGSNTTSQTS 345

RESULT 10

C42360

cellulase (EC 3.2.1.4) E5 precursor - *Thermomonospora fusca*

N:Alternate names: endo-1,4-beta-glucanase
 C:Species: *Thermomonospora fusca*

C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 20-Mar-1998
 C:Accession: C42360

R:Iao, G.; Ghangas, G.S.; Jung, E.D.; Wallson, D.B.
 J. Bacteriol. 173, 3397-3407, 1991

A:Title: DNA sequences of three beta-1,4-endoglucanase genes from *Thermomonospora fus*
 A:Reference number: A42360; MUID:91583320

A:Accession: C42360
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-466 <LAD>

A:Cross-references: GB:I01577; NID:9154693; PID:9154694
 C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Superfamily: bacterial cellulose-binding domain homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:38-137/Domain: bacterial cellulose-binding domain homology <BCB>

Query Match 10.8%; Score 176.5; DB 2; Length 466;
 Best Local Similarity 25.0%; Pred. No. 2.4e-06;
 Matches 78; Conservative 48; Mismatches 125; Indels 61; Gaps 17;

QY 8 VSGTTLVDANGNPFVNRGIN-HGHAWYKQAT-TAIEGIA-NTGANTVRI--VLSDDGGQW 62
 Db 173 VCGTQLCDEHNGVQLRGSMSTHGIOWFDECLTDSIDALAYDWKADIRLSMTYQEDGYE 232
 QY 63 TKQDIIHVR--NLISLAEDNHLVAPEVHDATGYDSIASLNRADVWIEMRSALIKEDT 120
 Db 233 TNPRGFTRIDQILDNAATAGLVYVDMWHLTPGDDPHYNTDRKTFEAFEAQRAHNAKTN- 291
 QY 121 VIINIANEWGSEGDAMAD--GY-KQAIPLRLNAGLHNTLWVDAAGWQGPQSHIDYGR 177

```

Db 292 VLXEIANE-----PNCVSNASIKSYAEVPIVIRORPDSVLIITGRGWSLGSV-EGSGP 346
QY 178 EYFNADP--QRTMFSIMHYEYAGNAAVPRINIDRYLNODLAL---VIGEGHRTND 232
Db 347 AEIAPNPVNASIMAFHEFY-----AASHRNYINALAEASELPFVFTEGETETTYTD 400
QY 233 -----VDEATIMSSEORGVGLAWSWK-----GNGPEWEYLDLSDNM 270
Db 401 GANDROMADR--YIDLMEARKIGWTKWYSDDFRSGAVRQPGTCASGSP-----W 448
QY 271 AGNNLTAMGNTI 282
Db 449 SCSXKASGOW 460

RESULT 11
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Uozumi, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NAE>
A:Cross-references: GB:M28332; MID:g142670; PIDN:AAA22307.1; PID:g142671
A:Experimental source: strain IFO3034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 10.7%; Score 174.5; DB 2; Length 499;
Best Local Similarity 21.1%; Pred. No. 3.8e-06;
Matches 70; Conservative 66; Mismatches 123; Indels 73; Gaps 15;

QY 3 NSGFVYSGTTLTDANGNPVNRGI-NHGAWYKDDQATTAIEGIANGVNTVRLVSDGQ 61
Db 39 NGOLSIKGTQLVNRDGAQVQLNGISGGLQWTD-----FYNKDSK 80
QY 62 WTKDD-----IHT-----YANLISLAEDNHLVAVPEVDATGYDIA 98
Db 81 WTRDWMGITVFRAAVTLDGGYIDNPVSKYKEAVEAKELGIYIIMHILNDGNPQ 140
QY 99 SINRAVDVIERKALIGEDTVIINIANEFG--SWEGD--AMADGYKQALPRLNAGL 154
Db 141 HREKAKDFEKE--SLYGMTPNVYIEIANEPNGDVMMKDIRPYAE--EVISVIRKNDP 196
QY 155 NHTLVADAAGWQFQSHIDYGRVFNADPO--RRTMFSIMHYEYAGNAAVPRINIDR 211
Db 197 DIIITV--GTTWSQDVND-----AADQLKANAVMTALHF--YAGTHGSLRDKANY 245
QY 212 VNODLALVIGEGHRTND-----VDEA-TIMSSEORGVGLAWSWKNGPEWEYL-- 264
Db 246 ALSKGAPITVETMGTSDAAGNGVFLDQSRRLNLTLSNINISVNMNLSDKDESSALAP 305
QY 265 --DLSNDNAGNNLTAMGNTIVNGPYGLRSTR 294
Db 306 GASKTGGMPLTDLTASGTFVRNLIANKDSTK 337

RESULT 12
S54744
cellulase (EC 3.2.1.4) Celv1 precursor - Erwinia carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwinia carotovora
A:Variety: SCC 3193

```

```

C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S54744; S44996
R:Maee, A.; Heikinhelmo, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwinia carotovora subspecies carotovora S54
A:Reference number: S54744; MUID:95231512
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <NAE>
A:Cross-references: EXBL:X79241; NID:g493492; PIDN:CAAS5823.1; PID:g493493
A:Genetics:
A:Gene: celv1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-504/Product: cellulase #status predicted <MAT>

Query Match 10.7%; Score 174.5; DB 2; Length 504;
Best Local Similarity 23.7%; Pred. No. 3.9e-06;
Matches 73; Conservative 43; Mismatches 123; Indels 69; Gaps 14;

QY 13 LYDANGNPVNRGI-NHGAWYKDDQATTAIEGIANGVNTVRLVSDGQWTKDD-IHT 69
Db 48 LVDEGKRYQLNGISNGIQWGD-----YANKDSKWLTRDMGINV 89
QY 70 VNLSLAEDNHLVAVPEV-----HDATGYDSIASLNRADYVW 107
Db 90 FRYAVYTLAE-NGYIANPSLANKVKEAVAAAGLGYIITIDMTLSDNDPNTYKAQAKTF 148
QY 108 TEMRSALICEKTYVITINIANEFGS--WEGD--AMADGYKQALPRLNAGLHTLMDAA 163
Db 149 AAM-AGLGNSPNVIYIEIANEPNGSVTNWQIRPYA--LEVTDITIRSKDPNLLIVSG 204
QY 164 GWGQFPOSIDYGRVFNADPORTMFSIMHYEYAGNAAVPRINIDRYLNODLALVIG 223
Db 205 TWS---QDIHD--AADNQLPDPNTLYALHF--YAGTHGQFLRDRIDVAGSGAAIFVSE 256
QY 224 FGRHRTND-----VDEATIMSSEORGVGLAWSWKNGPEWEYL--DLSNDWAGNN 274
Db 257 WTSQASGSGGFLEDSQWIDFLNNRGISWYMWLSLDSKERSAALVAGASGSGWTEQN 316
QY 275 LTMGNTI 282
Db 317 LSTSGKFV 324

RESULT 13
A26874
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.B.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <RDB1>
A:Cross-references: GB:M6185; NID:g143007; PIDN:AAA22496.1; PID:g143008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <RDB2>
A:Experimental source: strain DLG
A:Note: The authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminus
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

```

[illegible]

A: residues: 10-506 (MOL)
 A: Cross-references: EMBL:Z29076; NID:g509266; PIDN:CAA82317.1; PID:g509267
 R: Seo, Y.S.; Lee, Y.H.; Park, D.H.; Kang, H.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:06:42 ; Search time 248.55 Seconds
(Without alignments)
76.597 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGYVSGTLLDANGNP.....TIVNGPYGLRSTRSLSTVF 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues
Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 1169 | 71.7 | 516 | 2 | JEO134 |
| 2 | 710 | 43.6 | 384 | 2 | T37232 |
| 3 | 692 | 42.5 | 363 | 2 | S30386 |
| 4 | 601 | 36.9 | 1331 | 2 | A48954 |
| 5 | 202 | 12.4 | 409 | 2 | B25156 |
| 6 | 200 | 12.3 | 488 | 2 | A25156 |
| 7 | 197 | 12.1 | 557 | 2 | JCS487 |
| 8 | 184.5 | 11.3 | 505 | 2 | JCS962 |
| 9 | 180 | 11.0 | 448 | 2 | A27631 |
| 10 | 176.5 | 10.8 | 466 | 2 | C42360 |
| 11 | 174.5 | 10.7 | 499 | 2 | A27198 |
| 12 | 174.5 | 10.7 | 504 | 2 | S54744 |
| 13 | 173.5 | 10.6 | 508 | 2 | A26874 |
| 14 | 162.5 | 10.0 | 486 | 2 | I40548 |
| 15 | 162.5 | 10.0 | 486 | 2 | G65953 |
| 16 | 157 | 9.6 | 783 | 2 | JCS467 |
| 17 | 152 | 9.3 | 800 | 2 | A29003 |
| 18 | 152 | 9.3 | 822 | 2 | JT0611 |
| 19 | 151.5 | 9.3 | 499 | 2 | JT0111 |
| 20 | 151.5 | 9.3 | 825 | 2 | JT0174 |
| 21 | 149.5 | 9.2 | 357 | 2 | PC4404 |
| 22 | 149.5 | 9.2 | 941 | 2 | S28043 |
| 23 | 147.5 | 9.0 | 429 | 2 | S29044 |
| 24 | 136.5 | 8.4 | 570 | 2 | S56132 |
| 25 | 131 | 8.0 | 814 | 1 | CZCLEM |
| 26 | 130 | 8.0 | 426 | 2 | A42649 |
| 27 | 127 | 7.8 | 32 | 2 | PC4278 |
| 28 | 122.5 | 7.5 | 915 | 2 | A4802 |
| 29 | 122.5 | 7.5 | 1039 | 2 | S02711 |

| | | | | | | |
|----|-------|-----|-----|---|--------|--------------------|
| 30 | 122 | 7.5 | 438 | 2 | A47702 | glucan 1,3-beta-gl |
| 31 | 117.5 | 7.2 | 428 | 2 | S03767 | cellulase (EC 3.2. |
| 32 | 116 | 7.1 | 611 | 2 | JC7177 | endoglucanase V (E |
| 33 | 115.5 | 7.1 | 747 | 2 | B47093 | cellulase (EC 3.2. |
| 34 | 114 | 7.0 | 419 | 2 | S72325 | glucan 1,3-beta-gl |
| 35 | 113.5 | 6.8 | 517 | 2 | I40798 | cellulase (EC 3.2. |
| 36 | 111.5 | 6.8 | 441 | 2 | A44815 | cellulase (EC 3.2. |
| 37 | 111 | 6.8 | 475 | 1 | CZCICA | cellulase (EC 3.2. |
| 38 | 110 | 6.7 | 26 | 2 | PC4285 | guar gum-degrading |
| 39 | 109.5 | 6.7 | 754 | 2 | T14877 | hypothetical prote |
| 40 | 107 | 6.6 | 430 | 2 | S53325 | endo-beta-1,6-gluc |
| 41 | 105 | 6.4 | 364 | 2 | S12017 | endoglucanase A - |
| 42 | 105 | 6.4 | 456 | 2 | T40276 | probable gas1 fam1 |
| 43 | 104 | 6.4 | 566 | 2 | A40589 | cellulase (EC 3.2. |
| 44 | 101.5 | 6.2 | 317 | 2 | B72216 | endoglucanase - Th |
| 45 | 100.5 | 6.2 | 748 | 2 | S19652 | celldextrinase C |

ALIGNMENTS

RESULT 1
JEO134
manan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N: Alternate names: endo-1,4-beta-mannanase
C: Species: Bacillus circulans
C: Date: 03-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C: Accession: JEO134
R: Osada, S.; Sako, Y.; Uchida, A.
Biosci. Biotechnol. Biochem. 62, 514-520, 1998
A: Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A: Reference number: JEO134; MUID:98233274
A: Accession: JEO134
A: Molecule type: mRNA
A: Residues: 1-516 <POS>
A: Cross-References: DDBJ:AB007123
A: Note: The authors translated the codon CAA for residue 259 as Asn, CAA for residue
A: Note: the translation of the nucleotide sequence 1294-1365 is not given in this p
C: Keywords: glycoprotein; glycosidase; hydrolase

| | | | | |
|-----------------------|------------------|------------------------------------------------------------|--------------------------------------|-------------|
| Query Match | 71.7% | Score 1169; | DB 2; | Length 516; |
| Best Local Similarity | 69.2% | Pred. NO. 1e-83; | | |
| Matches 207; | Conservative 48; | Mismatches 44; | Indels 0; | Gaps 0; |
| QY | 1 | NANSGYVSGTLLDANGNP | FVWRGNGINCHANYKQOATTAIGCIANTYRIVLSDG | 60 |
| DB | 33 | HASGFTVSGTLLDANGNP | FVWRGNGINCHANYKQOATTAIGCIANTYRIVLSDG | 92 |
| QY | 61 | QWTKDHTVRLISLAEDNHLVAVPEVDATGYDSIASLNRAVYTEMRSALGKEDT | 120 | |
| DB | 93 | KTTLDVNTVNNITLCEQNKILAVLEVDATGSSLSLDNAVYTWIGIKSALGKEDT | 152 | |
| QY | 121 | VITINANEWEGSMEDGAWADGYKQAIPLRNAGLNTHTLVDAAGGQFPOSTHIDYREVF | 180 | |
| DB | 153 | VITINANEWEGSMEDGAWADGYKQAIPLRNAGLNTHTLVDAAGGQFPOSTHIDYREVF | 212 | |
| QY | 181 | NADPQNTMFSTHMYEYAGNAGSOVRTINDRYLNDLALVIESEHRRNTGVDATIMS | 240 | |
| DB | 213 | NADPQNTMFSTHMYEYAGNAGSOVRTINDRYLNDLALVIESEHRRNTGVDATIMS | 272 | |
| QY | 241 | YSEORGVGLAWSKNGPMEYELDSNDWAGNNTLVAGNTIVNGPYGLRSTRSLSTVF | 299 | |
| DB | 273 | YSEORGVGLAWSKNGPMEYELDSNDWAGNNTLVAGNTIVNGPYGLRSTRSLSTVF | 331 | |

RESULT 2
T37232
secreted beta-mannosidase - Streptomyces coelicolor
C: Species: Streptomyces coelicolor
C: Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C: Accession: T37232
R: Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

This Page Blank (uspto)

[illegible]

C:Function: A:Description: cleaves the beta-1,4-mannosidic linkages in various beta-mannans
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 16.4%; Score 427; DB 2; Length 513;
Best Local Similarity 26.9%; Pred. No. 9.6e-22;
Matches 149; Conservative 76; Mismatches 215; Indels 114; Gaps 23;

QY 5 EKRIISITLILASSILFVSGSTANANSSRYVSGTLLYDANGNPFWM----- 53
Db 4 YKKVAFVMAIEFFSVLPTIIMSSENN-----GALISNNANOTTKNYSWLANLPN 55
QY 54 ----RGINHG-AMYKQDATTATEGIANTGANTVRLVS-DGQW-----TKDDIHFTVN- 102
Db 56 KSNKRVRVSHFEGCYSDSTLAWIKQCARLGTGMKPIGLSDCYKNNQOTRLVADQISYGCNQ 115
QY 103 -LISLAEDNHLVAV-----PEVHDATGYDSIASLNRADV-----WIEKRSALI 145
Db 116 ELINFNQGLVTVISVHMPNPGFHSGENYKTLPTSQFONLTNNHRTTEGRKWKMDIKWA 175
QY 146 GRED-----YVLIINANEMFGSNREGDAW-ADGYKALIPRLNAGLN-----H 136
Db 176 DGIDELQNNGVYVLFPRPLEHNGEW--FWMGAEIGNQDQTRANAYISAMRDMOTQFTH 232
QY 187 TLMVDAAGNGQFPOS-HDQGREVFNDPQRTMFSTHMEYAGNAGQVRTNIDRYLND 246
Db 233 EKKLNLIWVSPDYTRD---VTSYTPGANYVDIYALDSY-HPDHSILDQYRRIAD 288
QY 247 LALVIEFGHRLH-NGDVDEATIMSYEQ--RGVGLAMS--WK--GNQPEWEYIDLSN 298
Db 289 KPRPAFLEIGPESMASFYSNYIAIKOKYPRTYFFLAMNDKSPHNNGAV---DLFN 345
QY 299 DNAGNNLTANGNTIVNGPYGLRETSRLSTFYTGSGSDGTS--PTLLYDEGSMQGTGS 356
Db 346 D-----SWVNRGEIDYQSNPATLYPENNLTLSMSC 379
QY 357 SLS-GGPWATWEMSSKSHSLKADIOSNSOYLYHVIONTSLOQSRIOATVKHANMS 415
Db 380 EFTDGGARLVSNENSSANGTOSLKADYVLGNNSYH-LQTYNRNLSFKNLEINVSWSGN 438
QY 416 VGNGTARLVYKTGHGYTWYSGSFVPINGSGTSLDLSNVONLSQVREICVQFOSASD 475
Db 439 VSGGMARARVYKIGSAMRWNAEGFCQFAGKRTALSIDLTKVNLHDVDEIGVEYKAPAN 498
QY 476 SSGQTSYIIDNVLY 489
Db 499 SNGKTAYIDHYTV 512

RESULT 6
JC5487
cellulase (EC 3.2.1.4) precursor - Clostridium cellulovorans
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium cellulovorans
C:Date: 07-Jul-1997 #sequence_revision 12-Sep-1997 #text_change 15-Oct-1999
R:Accession: J05487, PC4333
R:Shewite, S.A.; Ichi-Ishi, A.; Park, J.S.; Liu, C.; Malburg, L.M.; Dot, R.H.
A:Title: Characterization of engF, a gene for a non-cellulosomal Clostridium cellul
A:Reference number: J05487, M01D:97136706
A:Accession: J05487
A:Molecule type: DNA
A:Residues: 1-557 <SHE1>
A:Cross-references: GB:037056; NID:g1778708; PIDN:AA040891.1; PID:g1778709
A:Experimental source: strain ATCC 35296
A:Accession: PC4333
A:Molecule type: protein
A:Residues: 30-37 <SHE2>
C:Comment: This enzyme plays a secondary role in cellulose degradation.
C:Genetics:
A:Gene: engF
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 17:32:55 ; Search time 362.48 Seconds

(without alignments)
77.279 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630

Sequence: 1 NANSGRFVSGTLLYDANGNP.....TIVNGPYGLRNSRLSTVPT 300

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------------|
| 1 | 1172 | 71.9 | 516 | 2 | 066185 bacillus ci |
| 2 | 797 | 48.9 | 395 | 2 | 069347 vibrio sp. |
| 3 | 710 | 43.6 | 384 | 2 | 086599 streptomyces |
| 4 | 634 | 38.9 | 279 | 2 | 092F13 thermomicrobium |
| 5 | 614 | 37.7 | 290 | 2 | 092F13 thermomicrobium |
| 6 | 536.5 | 32.9 | 327 | 2 | 092F13 thermomicrobium |
| 7 | 226 | 13.9 | 1097 | 2 | 092A17 streptomyces |
| 8 | 197 | 12.1 | 557 | 2 | P94622 anaerobic t |
| 9 | 191 | 11.7 | 389 | 2 | P94622 anaerobic t |
| 10 | 176.5 | 10.8 | 501 | 2 | 083012 bacillus sp |
| 11 | 174.5 | 10.5 | 499 | 2 | 045532 bacillus su |
| 12 | 170.5 | 10.2 | 930 | 2 | 059290 clostridium |
| 13 | 166.5 | 10.0 | 387 | 2 | 031029 erwinia car |
| 14 | 162.5 | 10.0 | 486 | 2 | 045430 bacillus sp |
| 15 | 161.5 | 9.9 | 499 | 2 | 052731 bacillus sp |
| 16 | 160.5 | 9.8 | 481 | 2 | 066064 actinomyces |
| 17 | 159.5 | 9.8 | 635 | 2 | 066065 fibrobacter |
| 18 | 157 | 9.6 | 783 | 2 | 045554 bacillus sp |
| 19 | 154.5 | 9.5 | 749 | 2 | 059154 anaerocellu |

| | | | | | | |
|----|-------|-----|------|---|--------|---------------------|
| 20 | 152 | 9.3 | 821 | 2 | 059241 | 059241 bacillus sp |
| 21 | 140 | 8.6 | 278 | 2 | 09R56 | 09R56 streptomyces |
| 22 | 136.5 | 8.4 | 570 | 2 | 059665 | 059665 pseudomonas |
| 23 | 129 | 7.9 | 478 | 5 | 016028 | 016028 globodera r |
| 24 | 126 | 7.7 | 319 | 5 | 018454 | 018454 heterodera |
| 25 | 126 | 7.7 | 438 | 3 | 09UR18 | 09UR18 candida alb |
| 26 | 126 | 7.7 | 476 | 5 | 018453 | 018453 heterodera |
| 27 | 125 | 7.7 | 395 | 5 | 0906M4 | 0906M4 globodera t |
| 28 | 123 | 7.5 | 319 | 5 | 061595 | 061595 heterodera |
| 29 | 123 | 7.5 | 319 | 5 | 077449 | 077449 heterodera |
| 30 | 121.5 | 7.5 | 1000 | 2 | 024820 | 024820 thermophilu |
| 31 | 117 | 7.2 | 391 | 5 | 077094 | 077094 globodera r |
| 32 | 117 | 7.2 | 392 | 2 | 044078 | 044078 globodera r |
| 33 | 116.5 | 7.1 | 332 | 2 | 060054 | 060054 unidentified |
| 34 | 116 | 7.1 | 426 | 2 | 09REW0 | 09REW0 erwinia chr |
| 35 | 115 | 7.1 | 363 | 2 | 007652 | 007652 celivibrio |
| 36 | 114 | 7.0 | 419 | 3 | 012539 | 012539 agardicus bl |
| 37 | 114 | 7.0 | 470 | 5 | 0906M5 | 0906M5 globodera t |
| 38 | 113 | 6.9 | 910 | 3 | P87211 | P87211 orpionmyces |
| 39 | 112.5 | 6.9 | 621 | 2 | 007653 | 007653 celivibrio |
| 40 | 111.5 | 6.8 | 506 | 5 | 090A57 | 090A57 meloidogyne |
| 41 | 110 | 6.7 | 435 | 3 | 012626 | 012626 pichia angu |
| 42 | 109.5 | 6.7 | 754 | 2 | 085318 | 085318 salmonella |
| 43 | 107 | 6.6 | 430 | 3 | 012712 | 012712 trichoderma |
| 44 | 106 | 6.5 | 325 | 3 | 09Y8H6 | 09Y8H6 emericella |
| 45 | 105 | 6.4 | 456 | 3 | 09Y7Y7 | 09Y7Y7 schizosacch |

ALIGNMENTS

| RESULT | 1 | PRELIMINARY: | PRT: | 516 AA. |
|-----------------------|---------------------------------------------------------------------|--------------|------|--------------------|
| 066185 | 066185 | | | |
| AC | 066185 | | | |
| DT | 01-AUG-1998 (TRENBLER, 07, Created) | | | |
| DT | 01-AUG-1998 (TRENBLER, 07, Last sequence update) | | | |
| DT | 01-MAY-2000 (TRENBLER, 13, Last annotation update) | | | |
| DE | MANNANASE. | | | |
| OS | Bacillus circulans. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| OC | Bacillus/Staphylococcus group; Bacillus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE; 98233274. | | | |
| RA | Yoshida S., Sako Y., Uchida A.; | | | |
| RT | Cloning, sequence analysis, and expression in Escherichia coli of a | | | |
| RT | gene coding for an enzyme from Bacillus circulans K-1 that degrades | | | |
| RT | guar gum." | | | |
| RL | Biosci. Biotechnol. Biochem. 62:514-520(1998). | | | |
| DR | EMBL; AB007123; BAA25878.1; - | | | |
| DR | INTERPRO; IPR001547; - | | | |
| DR | PFAM; PF00150; cellulase; 1. | | | |
| SO | SEQUENCE 516 AA; 55245 MW; DBE5A48842AE8062 CRC64; | | | |
| Query Match | 71.9%; Score 1172; DB 2; Length 516; | | | |
| Best Local Similarity | 69.2%; Pred. No. 2e-84; | | | |
| Matches | 207; Conservative 50; Mismatches 42; Indels 0; Gaps 0; | | | |
| QY | 1 NANSGRFVSGTLLYDANGNPVFGNGINHGAMVYDQATTAIEGIANGCANTVRIYLSGG 60 | | | 059241 bacillus sp |
| DB | 33 HAASGFYSGTLLYDANGNPVFGNGINHGAMVYDQATTAIEGIANGCANTVRIYLSGG 92 | | | 09R56 streptomyces |
| QY | 61 QMTDDITFVNLISLADNHLVAVPEHDTGYDSISLRAYDWTGEMSAIGREDT 120 | | | 059665 pseudomonas |
| DB | 93 KMTLDVATVNNITLCEQNKILAVLEHDTGSDLSLDDNNAVYWGISALGKDR 152 | | | 016028 globodera r |
| QY | 121 VIINANEMFGSWGDAWDGYKQAIPLRNAGLNHTLMDVDAWGQFPOSIDHYGREVF 180 | | | 018454 heterodera |
| DB | 153 VIINANEMFGSWGDAWDGYKQAIPLRNAGLNHTLMDVDAWGQFPOSIDHYGREVF 212 | | | 09UR18 candida alb |
| QY | 181 NADFORNTMESIHETVYAGNAGVATINIDRYLNMOLALVIGFGHRTNGDVDAATIMS 240 | | | 018453 heterodera |

DB 213 NADPLKNTVFISIHMEYAGNASTYKSNIDCVLNKLNALLIGERGAGGTNDVDVEATLMS 272
QY 241 YSEORGVCWGLAMSKNGPEWEYDLSDNMGNNLTANGNTIVNGPYGLRSTSLTVF 299
DB 273 YSOERGVCWGLAMSKNGSNDSALATLDMTNDMAGNSLTSGNTVNGSNCKATSTVLSGIF 331

RESULT 2
ID 069347 PRELIMINARY: PRT: 395 AA.
AC 069347
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-1,4-MANNANASE.
GN MANA.
OS Vibrio sp.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA-138;
RA Tamaru Y., Arai T., Morishita T., Kimura T., Sakka K., Ohmura K.;
RL J. Ferment. Bioproc. 83:201-205(1997).
DR EMBL; D86329; BAA25188.1; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002883; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF02013; CBD.5; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 395 AA: 43098 MW; A3E30B3A878C5DA1 CRC64;

Query Match 48.9%; Score 797; DB 2; Length 395;
Best Local Similarity 50.8%; Pred. No. 4.4e-55;
Matches 155; Conservative 55; Mismatches 87; Indels 8; Gaps 5;

QY 2 ANSGYVSGTLYDANGNPFVNRGINHGHWYKQDATTAEIGANTGANTVRYLSGQ 61
DB 19 AAGGIVHSNGVLYEANGSFKIRGINHATWYTDRLSVLSGIATGANTVRYVLSNGR 78
QY 62 WTKODIHVYRLISLAEDNHLVAVPEVHDATGY--DSIASNRAVDYIEMRSALIGKE 118
DB 79 WTKNVSDVTNINIAKANNLTAILEVHDPTGYGESSAASLSDADTYIELAKNELIGOE 138
QY 119 DTVIINIANEFWSWEGDAMADGYKQALPRLNAGLNHTIYNVDAAGWGQ-FPOSINDYGR 176
DB 139 DVIINILGNEPFGNNNDVAVVNDHVSATQRLRSAGINHTIYNVDAAPNMGODKGFMLNNA 198
QY 177 REVFNADPQRNMFESIHMEYAGNASTYKSNIDCVLNKLNALLIGERGAGGTNDVDVEAT 235
DB 199 QVFNMSDPLNITFVHMEYVSYNS--VNDYISSFTNGVLVYGEFASVTHKADVDE 256
QY 236 ATINSYSEORGVCWGLAMSKNGPEWEYDLSDNMGNNLTANGNTIVNGPYGLRSTSL 295
DB 257 GSIMERSETLSGTYGMSNGSDTTSDLDYNNMDNNSYSTWGNVLINGNGISTSTL 316
QY 296 STVFT 300
DB 317 ATVFT 321
RESULT 3
ID 086599 PRELIMINARY: PRT: 384 AA.
AC 086599
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE SECRETED BETA-MANNOSIDASE.
GN MANA2.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Randalream M.A.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RT Kinash H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031514; CA20610.1; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002883; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF02013; CBD.5; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN.1.
SQ SEQUENCE 384 AA; 39847 MW; B8AD1583FD3BD11 CRC64;

Query Match 43.6%; Score 710; DB 2; Length 384;
Best Local Similarity 46.7%; Pred. No. 3e-48;
Matches 141; Conservative 53; Mismatches 102; Indels 6; Gaps 4;

QY 2 ANSGYVSGTLYDANGNPFVNRGINHGHWYKQDATTAEIGANTGANTVRYLSGQ 61
DB 37 AAGGIVHSNGVLYEANGSFKIRGINHATWYTDRLSVLSGIATGANTVRYVLSNGR 95
QY 62 WTKODIHVYRLISLAEDNHLVAVPEVHDATGY--DSIASNRAVDYIEMRSALIGKE 118
DB 96 WTKNSASEVSLAQCKANKVYCLEVHDPTGYEDGATSLSDDADYVWSYKALSQOE 155
QY 119 DTVIINIANEFWSWEGDAMADGYKQALPRLNAGLNHTIYNVDAAGWGQ-FPOSINDYGR 177
DB 156 DVIYVNGNEPFGNTATAMDATKSAIGKRGAGLDHALVADAPNMGODKGFMLNNA 215
QY 178 EVFNADPQRNMFESIHMEYAGNASTYKSNIDCVLNKLNALLIGERGAGGTNDVDVEAT 237
DB 216 SVFASDPLNITFVHMEYVSYNS--VNDYISSFTNGVLVYGEFASVTHKADVDE 274
QY 238 IMASSEORGVCWGLAMSKNGPEWEYDLSDNMGNNLTANGNTIVNGPYGLRSTSL 297
DB 275 IMATROSLGVLYLMSNGSGNGVYEDLDVNGFDPNSITSGNRIFGSNGIATSTRTAT 334
QY 298 VF 299
DB 335 VY 336
RESULT 4
ID 092F13 PRELIMINARY: PRT: 279 AA.
AC 092F13
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-MANNANASE (EC 3.2.1.78) (FRAGMENT).
GN MAN.
OS Thermomonaspora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KW3;

| Query Match | 38.9% | Score 634; | DB 2; | Length 279; |
|-------------------------------------------------------------------------|---------------------------------------------------------------|--------------------|-------------------------|-------------|
| Best Local Similarity | 44.48; | Pred. NO. 1.8e-42; | | |
| Matches 128; | Conservative 56; | Mismatches 82; | Indels 22; | Gaps 6; |
| QY 5 | GFVYSGLTYLDXANPFWVARGINHGHWYTKDOATTAIEGIANTGANTVATVIVSDGGQWTK | 64 | | |
| DB 1 | GLHWKNGRLRYEANGGEFIRGVSHPHNMY-PQHTQAFADIKSHGANTVAVVLSNGVRMSK | 59 | | |
| QY 65 | DIHRYRNLISLAEDNHLVAVPEVDATGY---DSIASLNRAVDVYIEKRSALLIGEDTV | 121 | | |
| DB 60 | NGPSDVANYISLCKONRLICMLVEHDTGTGSGOSGASTLDOAVDWIIEKSTVLOGEEDTV | 119 | | |
| QY 122 | ININANEMF-----GSWEGDAMADQKQAIPLPRNAGNHTLMTVAADGCGQ-FPOS | 171 | | |
| DB 120 | LINIGNEPYGNDASVYVAGAMDTSA-----AIQRLRAGAEHRTLVYDAPRMGQDWTWT | 172 | | |
| QY 172 | HDYGREVFNADPQRTNMFSSIMETETAGGNASQVRTNIDRVINQDLAVYIGFGRHRTT | 231 | | |
| DB 173 | MKNNDADQVYASAPGNTNVSIFHYMG-VYSQASTITSYLEHFVAGIPLIIGFGRHSDG | 231 | | |
| QY 232 | DVDEATINSYSGRQGVGLWMSKMGKPMWEYLDLSNDMAGNNLTRMG | 279 | | |
| DB 232 | NDEDTIMAEARLKLGTIGMSWNGGGEVETLMDYTNFDPGNTLSPWG | 279 | | |
| RESULT 5 | | | | |
| Q9RFXS | PRELIMINARY: | PRT: | 930 AA. | |
| AC 09REXS | | | | |
| DT 01-MAY-2000 | (TREMBLrel. 13, Created) | | | |
| DT 01-MAY-2000 | (TREMBLrel. 13, last sequence update) | | | |
| DT 01-JUN-2000 | (TREMBLrel. 14, last annotation update) | | | |
| DE MULTIDOMAIN | BETA-1,4-MANNANASE PRECURSOR. | | | |
| GN MANA. | | | | |
| OS Caldicellibacillus cellulovorans | | | | |
| OC Bacteria; Firmicutes; Bacillus/Clostridium group; | | | | |
| OC Bacillus/Staphylococcus group; Caldicellibacillus. | | | | |
| RN [1] | | | | |
| RP SEQUENCE FROM N.A. | | | | |
| RA Suna A., Gibbs M.D., Bergquist P.L.; | | | | |
| RT "A novel multidomain beta-1,4-mannanase gene from Caldicellibacillus | | | | |
| RT cellulovorans and action of the recombinant enzyme on kraft pulp."; | | | | |
| RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases. | | | | |
| DR EMBL, AF163837; AAF22274.1; -. | | | | |
| DR INTERPRO; IPR001547; -. | | | | |
| DR INTERPRO; IPR001956; -. | | | | |
| DR | | | | |
| RA Hilge M. Gloor S.M., Winterhalter K., Zimmermann W., Piontek K.; | | | | |
| RT "Crystallization and preliminary crystallographic analysis of two | | | | |
| RT beta-mannanase isoforms from Thermomonospora fusca KW3."; | | | | |
| RL Acta Crystallogr. D 52:1224-1225(1996). | | | | |
| RN [2] | | | | |
| RP SEQUENCE FROM N.A. | | | | |
| RC STRAIN-KW3; | | | | |
| RX MEDLINE; 99036862. | | | | |
| RA Hilge M., Gloor S.M., Rypniewski W., Sauer O., Heigtmann T.D., | | | | |
| RT Zimmermann W., Winterhalter K., Piontek K.; | | | | |
| RT "High-resolution native and complex structures of thermostable beta- | | | | |
| RT mannanase from Thermomonospora fusca - substrate specificity in | | | | |
| RL glycosyl hydrolase family 5."; | | | | |
| RL Structure 6:1433-1444(1998). | | | | |
| CC -I- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC | | | | |
| CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND | | | | |
| CC GALACTOGLUCOMANNANS. | | | | |
| CC EMBL; AJ006227; CAA06924.1; -. | | | | |
| DR INTERPRO; IPR001547; -. | | | | |
| DR PFAM; PF00150; cellulase; 1. | | | | |
| DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1. | | | | |
| DR Hydrolase; Glycosidase. | | | | |
| FT NON_TER | 1 | | | |
| FT NON_TER | 279 | | | |
| FT NON_TER | 279 | | | |
| SQ SEQUENCE | 279 AA; | 30657 MW; | 98A2850FAE013DB0 CRC64; | |

```

DR INTERPRO: IP002965; -
DR Pfam: PF00150: cellulase, 1.
DR Pfam: PF00942; CBD_3; 2.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Signal.
FT CHAIN 1 33 POTENTIAL.
FT SEQUENCE 34 930 MULTIDOMAIN BETA-1,4-MANNANASE.
SQ SEQUENCE 930 AA; 101576 MW; 0086638D54D1A2CC CRC64;

Query Match 37.7%; Score 614; DB 2; Length 930;
Best Local Similarity 46.9%; Pred. No. 3,9e-40;
Matches 122; Conservative 52; Mismatches 80; Indels 6; Gaps

QY 25 GINHEHANYTKDQATTAIGTANTGANYRYILVSGGCGTKDDHITVKNLISLAEQ-NHLY 83
DB 486 GINHPHANYRRLRSSLOGISMGWANNARYILVSGCRKTKPASEVADITISQARTLQYRA 545
QY 84 AYPEHDAATGY-DSEA-SLNRAVDYWIEMRSALIGKEDYVILINIAEMWESGEGDAMAD 140
DB 546 VLYEVHDTGTGEGDAACASMTTANYWTELKNAVLAGENFIVYVIGNEPNNNYQWMT 6050
QY 141 GYKQALPRLNAGLNTMLVDAAGMGQ-FPSIHQGYEYFNADPQNTMFSIMYEYAG 1999
DB 606 DTRNAYVALRNGINNTIMVDAIPWVGQDWSFTMDNPTIENADPQRNLVFSIMYG-VY 6644
QY 200 GNASOVNRINIPVNLQDIALVIGEGRHRTNGDQVDEATINSYSPQRQVGLMAWKGNGP 2599
DB 665 DTRAEVQSTIESFYNRGLPLVYIGEGHMSDGPDEQALVQYAKQVNLGLFGWSGNG 7244
QY 260 EWEYLDLSNDVAGNLTWANG 279
DB 725 GVEYEDVMTNFMNANGPTWANG 744

RESULT 6
Q9RJ15 PRELIMINARY; PRT; 327 AA.
AC Q9RJ15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
GN MANA.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parthali J., Barrell B.G., Rajandream M.A.;
RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinash H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
DR EMBL: AL121746; CAB57406.1; -
DR INTERPRO: IP001547; -
DR Pfam: PF00150: cellulase; 1.
DR NON_TER 327
SQ SEQUENCE 327 AA; 34772 MW; A21B0068C2C0975D CRC64;

```



```

Db      215 STANGETEONTPTVDEKRSALVTYGGDLNADBSAKRTIFAFINYN-EGDIQKVEYDIDR 273
Oy      212 VIANODLALVIGEGHRTNGVDDEATIMSYSE---ORGYGLWASR-----KGN 257
Db      274 ANMAGLVFVEEYEGKQDS--DAAEGYKSGGLQVNMKKGAGRIYMWMDGYDLDTISGTR 331
Oy      258 GPWEYIDLDNDMAGNNLTAMGNTIYNGPYGLRET 292
Db      332 GSGWE-INKTDGSKPTVLSWVGDKINDNNGIIFT 365

RESULT      8
P946622      PRELIMINARY;      PRT;      557 AA.
AC      01-MAY-1997 (TIMBLrel. 03, Created)
DT      01-MAY-1997 (TIMBLrel. 03, last sequence update)
DT      01-MAY-2000 (TIMBLrel. 13, last annotation update)
DE      ENDO-1,4-BETA GLUCANASE ENGAF (EC 3.2.1.4) (CELLULASE) (ENDOGALACTANASE;
GN      (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
OS      Clostridium cellulovorans.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC      Clostridium.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDJING: 97136706.
RA      Shevella S.A., Ichit-Ishi A., Park J.S., Liu C., Malburg L.M.,
RA      Dol R.H.;
RA      "Characterization of engF, a gene for a non-cellulosomal Clostridium
RT      cellulovorans endoglucanase."
RL      gene 192:163-167(1996).
CC      -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC      LINKAGES IN CELLULOSE.
CC      EMBL: U37056; AAA80891.1; -.
DR      HSSP: 085465; ZACH.
DR      INTERPRO: IPR001547; -.
DR      Pfam: PF00150; cellulase_1.
DR      PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW      Hydroxylase; Glycosidase.
SQ      SEQUENCE 557 AA; 60131 MW; D186CE88EB504EED CRC64;

```

```

RESULT 9
ID 059232 PRELIMINARY; PRT; 389 AA.
AC 059232;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE ENDO-BETA-1,4-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULOSE)
DE (ENDOGALACTAMINASE) (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA Sanchez Torres J.;
RC STRAIN-186-1;
RL Thesis (1994), Universidad de Salamanca, Spain.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
DR EMBL; Z3876; CA83942.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPRO01547; -.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 389 ENDO-1,4-BETA-GLUCANASE.
SQ SEQUENCE 389 AA; 43688 MW; 9111397485609B5F CRC64;

Query Match 11.7%; Score 191; DB 2; Length 389;
Best Local Similarity 25.1%; Pred. No. 2.2e-07;
Matches 70; Conservative 45; Mismatches 102; Indels 62; Gaps 13;

QY 8 VSGTLYDANGNPFVARGI-NHGAWYKQDATTAEIGANTGANTVRIYLSDDGQTKDD 66
DB 40 IINGELVNRGRQVQLKGSISHLQWYGD-----QFNYEEMKLRDD 81
QY 67 -----IHT-----VRNLISLAEDNHLVAPEVDATGYSIASLRA 103
DB 82 WGITVFRAMYTSSGGYIDDPVKEKVEAVEAIDLDIYIIDHILSDNPNIYKEEA 141
QY 104 VDYVIEKMSALIGKEDYVIINIANEFSGWEDDANADGRK---QAIPLRNAGLNHTLM 159
DB 142 KQFEEM-SELYGDPYNYETIETIANENGSG--DYTWDRKIPKPAEETIPYIRNNDPNNII 198
QY 160 VDAAGMGQPOSIDHGRVFNADPORNTMFSIMHYETAGNAGSOVRTNIDRYLNDLAL 219
DB 199 V---GTGWSQVH-HAADNQLADP--NYMIAFHF--YAGTGGONLRDQVDTALDGAH 250
QY 220 VIGFGHRTNGD---VDEA-TMYSSEQRGVGLAWS 253
DB 251 FVSEWGTSAATGDDGVFLEDAEWMIDFMDERNLSTANWS 289

RESULT 10
ID 083012 PRELIMINARY; PRT; 501 AA.
AC 083012;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE CELLULOSE.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA Khamonuch C., Ooi T., Kinsoshita S.;
RC STRAIN-5H;
RT "Cloning and nucleotide sequence of beta-mannanase and cellulase gene
from Bacillus sp. 5H.";
RL Submitted (JUL-1998) to the EMBL/Genbank/DBD databases.

```

```

DR EMBL; AB016164; BA31712.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPRO01547; -.
DR INTERPRO; IPRO01956; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PRODOM; PD001947; -. 1.
SQ SEQUENCE 501 AA; 55422 MW; F76D260A901E2DID CRC64;

```

```

Query Match 10.8%; Score 176.5; DB 2; Length 501;
Best Local Similarity 21.1%; Pred. No. 4.4e-06;
Matches 70; Conservative 67; Mismatches 122; Indels 73; Gaps 15;

QY 3 NSGFVSGTLYDANGNPFVARGI-NHGAWYKQDATTAEIGANTGANTVRIYLSDDGQ 61
DB 39 NGDLSTKGTGLINQNKANQVLKGISISHLQWYGD-----FVKKDSLK 80
QY 62 WTRKD-----IHT-----VRNLISLAEDNHLVAPEVDATGYSIA 98
DB 81 WLRDDGITVFRAMYTADGGYIDNPSRKRYKAVEAKELGIYIIDHILNDGNPNQ 140
QY 99 SLNRADVIERKMSALIGKEDYVIINIANEFSGWED--AMADGYQALPLRNAGL 154
DB 141 NKEKKEFFKEKES-LYGTWPVNYEIEENEPGVDWKKRDIPTVAE--EVSIVIRKNDP 196
QY 155 NHTLWDAAGMGQPOSIDHGRVFNADPQ---RTMFSIMHYETAGNAGSOVRTNIDR 211
DB 197 DNTIIV---GTGWSQVH-----AADQKLDANVMYALHF--YAGTGGOSLNDKANY 245
QY 212 VINQDLALVIGFGHRTNGD---VDEA-TMYSSEQRGVGLAWSKNGPWEYTL-- 264
DB 246 ALSKGAPIFTVETWGSIDASNGGVFLDQSGREMLNYLDSKISVMVNNLSDKQESSALKP 305
QY 265 --DLSNDAGNNLTAMGNTIVNGPYGLRSTR 294
DB 306 GASKTGGMPLTDLTASGTFVRNIFGNKDSRK 337

```

```

RESULT 11
ID 045532 PRELIMINARY; PRT; 499 AA.
AC 045532;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE CELLULOSE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura A., Uozumi T., Teruhiko B.;
RT "Nucleotide sequence of a cellulase gene of Bacillus subtilis.";
RL Eur. J. Biochem. 164:317-320(1987).
DR EMBL; M28332; AAA22307.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPRO01956; -.
DR INTERPRO; IPRO01547; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00942; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PRODOM; PD001947; -. 1.
SQ SEQUENCE 499 AA; 55075 MW; DA24AB9E63B94D23 CRC64;

```

```

Query Match 10.7%; Score 174.5; DB 2; Length 499;
Best Local Similarity 21.1%; Pred. No. 6.2e-06;
Matches 70; Conservative 66; Mismatches 123; Indels 73; Gaps 15;

QY 3 NSGFVSGTLYDANGNPFVARGI-NHGAWYKQDATTAEIGANTGANTVRIYLSDDGQ 61

```

```

Db      39 NQSLIKGTQVNVNBDKAKVQJLKGISBGLQWYGD-----FYNKDSLK 80
QY      62 WKKD-----IHT-----VRNLISLAEENHLVAVPEVDATGYDSIA 98
Db      81 WLRDWMGITVPRAMITTDGCIIDNPVKNKYKEAVEAKELGIYVIDMHLNDGNPNQ 140
QY      99 SLNRADVWIEMRSALIGKEDTVIINIANEFQ--SWEGD--AMADGKQALPRLRNGL 154
Db      141 HKEKADPFKEM--SLYGTNPVYIYEIANEPGDVNMKRDIPYAE---EYISYIRKNDP 196
QY      155 NRTIAMDAGNGQFPOSIDYGRFVNADPO---RRTMSIMTYEYAGNNSOVRTNIDR 211
Db      197 DNTIIV--GTGTWSDVDND-----AADQLKDNVYALHF--YAGTHGQSLDXYNY 245
QY      212 VLNODLALVIGERGHRTNGD---VDEA--TMSYSEORGWGLAMSKNGPEPEYL-- 264
Db      246 ALSKGAPITVTEMTGSDASNGGVFLDQSRWMLNLYDSKNISWYNNLSDKQESSALKP 305
QY      265 --DLSNDMAGNNTIANGNTIVNGPYGLRETSR 294
Db      306 GASKTGWPLTDLTASGTVRENILNKDST 337

```

```

RESULT 12
Q59290      PRELIMINARY;      PRT;      930 AA.
AC 059290;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE BETA-1,4-ENDOGLUCANASE (EC 3.2.1.4).
GN CELA.
OS Clostridium jostii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group: Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RA Fujino T., Fujino E., Karita S., Ohmura K.;
RT "Revised sequence of cels gene encoding endoglucanase (Eg)-1 from
RL Clostridium jostii";
DR EMBL; D85326; BAI2826.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001119; -.
DR INTERPRO; IPR001547; -.
DR PRAM; PF00150; cellulase; 1.
DR PRAM; PF00395; SLH; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR Hydrolase; glycosidase.
SQ SEQUENCE 930 AA; 101726 MW; 0E8AE8754D084448 CRC64;

```

Query Match 10.5%; Score 170.5; DB 2; Length 930;
 Best Local Similarity 26.0%; Pred. No. 3.2e-05;
 Matches 78; Conservative 45; Mismatches 106; Indels 71; Gaps 19;

```

QY      12 TLYDANGNPFVANGI-HGHANY-----KQATTAEGTANTGANTVRYL--SSGGQWTX 64
Db      61 TLCDKGNDFIQLRGSTHGLQWFPETINNAPALS--KDWGSNVIRLAMYABEGSKD 118
QY      65 DDIHTVRLN--ISLAEENHLVAVPEVDATGYDSIASLNR--AVDYIEMRSALIGKEDTV 121
Db      119 PELIKRVRIDGIDLATANDYIVVDHVTLPDPDNDVYKGMDFKEL--SQKYNNPDI 177
QY      122 IINIANEFQSWEG--DA--WA--DGYKQALP--LRNAGLHTIAMDAGNGQFPOSIH 173
Db      178 IYLANEPSPNDGVTNDAGNKAQVSYAEPIIKILRDSGNLNLIVGSPNNSQRP----- 233
QY      174 DYREVPNADPQNTYFSLHMYEYAGNNSOVRTNIDRYLNDLALVIGFGRHTNGDV 233
Db      234 DLAAE--NPINDNNTAIVSFHYSGT-----HKSTIDST 264

```

```

QY      234 DEATIMS---YSEORGWGLAMSW-----KGNP-----EW-EYLDLSN-DMAGNLT 276
Db      265 DRGNIMSNARYALEHGVAVFCSEMGTSSEASGNNGPYLKEADEWLEFELANNISWINSLT 324

```

```

RESULT 13
Q31029      PRELIMINARY;      PRT;      387 AA.
ID 031029;
AC 031029;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
GN CELA.
OS Erwina carotovora subsp. carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwina.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT34;
RA Park Y.W., Lim S.T., Yun H.D.;
RL Mol. Cells 0:0-0(1997).
DR EMBL; AF025768; AAC02964.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001547; -.
DR PRAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR Signal; Hydrolase.
FT SIGNAL 1
FT CHAIN 32 387
FT SEQUENCE 387 AA; 42003 MW; 6302B3DF8AC2738B CRC64;

```

Query Match 10.2%; Score 166.5; DB 2; Length 387;
 Best Local Similarity 21.5%; Pred. No. 1.8e-05;
 Matches 68; Conservative 50; Mismatches 113; Indels 85; Gaps 13;

```

QY      13 LYDANGNPFVANGI-NHGANYKQDQATTAEGTANTGANTVRYLSGGQWTKDD----- 66
Db      48 LVDEGKRVQJLKGISBGLQWYGD-----YRKDKSMKWLDPDQWGINV 89
QY      67 -----IHTVRNLISLAEENHLVAVPEVDATGYDSIASLNRADVYI 108
Db      90 SEVAMITADGYSKPSIANKYKEAVEAKELGIYVIDMHLNDNPYIKKEAKTFFA 149
QY      109 EKRSALIGKEDTVIINIANEFQ--SWEGD--AMADGKQALPRLRNAGLHTIAMDAG 164
Db      150 EM-AGLIGNSPIVYIYEIANEPGVTWDEIRPYA---LEVETIRSKDPNLTIV--G 202
QY      165 WQFPOSIDYGRFVNADPQNTYFSLHMYEYAGNNSOVRTNIDRYLNDLALVIGEF 224
Db      203 TGTWSDIHD---AADNQLPDNTIYALHF--YAGTHGQFLRIRIDYQSGAMIFSEW 257
QY      225 GHRHTNGD---VDEATIMSSEORGWGLAMSKNGPEPEYLDLS----- 267
Db      258 GTSDASGNGPFLPESHWTYIDLINRG-----GSRVNSLSDKSEASALAPGASK 308
QY      268 -NDMAGNNTIANGNTI 282
Db      309 CGGWTBQNLASGKFV 324

```

```

RESULT 14
Q45430      PRELIMINARY;      PRT;      486 AA.
ID 045430;
AC 045430;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
GN CEL.
OS Bacillus sp.

```

| Query | Match | Best Local Similarity | Score | DB 2 | Length | 486 |
|---------|---------------------------------------------------------------|--------------------------------------------------------------|-------|------------|--------|--------------------|
| Matches | 66 | Conservative | 75 | Mismatches | 136 | Indels 37; Gaps 15 |
| DB | 3 | NSGFVSSGTTLYDANGNPFVARGI-NBHGAKTKDQAT-TAIGCIANTGANVY---RIYLS | 57 | | | |
| DB | 39 | NGOLSTKCTQVNRGRKAVOLKGISSHLOWGEYVNRKDSLMLDMDWGTTFVRFAMRYA | 98 | | | |
| DB | 58 | DGGQWTKDQI-HIVRNLSIAEDNHLVAPPEVHDATGVYSIASLNRRAVDWYTEMRSALTG | 116 | | | |
| DB | 99 | DGGITDNPSSVNRKRAVEAFAKELGIYITDHIILNDGNPNONKEKAEFFEREM-SSLTG | 157 | | | |
| DB | 117 | KEDVYIINIAEMFEG-SWEGD-AMADGYKQALPRLNAGNGLHMLAMDVAWGQFPOSI | 172 | | | |
| DB | 158 | NTPNVIYIAENPNDDVMMKRIKRYAE---EVIAYIRKNDPNIIIV---GTGIMSQVY | 211 | | | |
| DB | 173 | HDYGEVFNADPQ---RNTMFSIHMYETVAGNASQVRINIDRYLMDLALVIGFGHRT | 229 | | | |
| DB | 212 | ND-----AADDQLDADAVMYALHF--YAGTGHGFLDKRANALSKGAPIFVTEWGTSDA | 263 | | | |
| DB | 230 | NGD-----VDEA-TIMSYSEQRGVGLWAMSKNGNGEWEL-----DLSNDMAANNLTAMN | 280 | | | |
| DB | 264 | SGNGGVTFEDQSEWTKYIDSKTISVWNWMLSKQSSSALKPGASKGTGWRSLDLSASGT | 323 | | | |
| DB | 281 | TIYNGPIGLRETSR | 294 | | | |
| DB | 324 | FVRENILGTNDSTK | 337 | | | |
| RESULT | 15 | | | | | |
| ID | 052731 | | | | | |
| AC | 052731 | PRELIMINARY; | PRT; | 499 | AA. | |
| DT | 01-JUN-1998 | (TREMblrel. 06, Created) | | | | |
| DT | 01-JUN-1998 | (TREMblrel. 06, Last sequence update) | | | | |
| DT | 01-JUN-2000 | (TREMblrel. 14, Last annotation update) | | | | |
| DE | ENNO-B-1.4-GLUCANASE (EC 3.2.1.4) (CELLULASE) (ENOGGLUCANASE) | | | | | |
| DE | (ENDO-1.4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULASE). | | | | | |
| GN | CELS. | | | | | |
| OS | Bacillus sp. | | | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | | | |

CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Jung K.-H., Chun Y.-C., Lee J.-C., Kim J.-H., Yoon K.-H.;
 RL Biotechnol. Lett. 18:1077-1082(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Yoon K.-H., Jung K.H.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ENDOTHEROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 DR EMBL: AF045482; AAC02536.1; -.
 DR HSSP: 085465; 2A3H.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.
 DR PFM: PF00150; cellulase; 1.
 DR PFM: PF00942; CBD_3; 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM: PD001947; -1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 499 AA; 55193 MW; 990A18786CF738 CRC64;

[illegible]

Search completed: December 19, 2000, 17:32:58
Job time: 1525 sec

Wed Dec 20 15:36:52 2000

us-09-339-159-2_copy_31_330.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:04:10 ; Search time 226.76 Seconds

(without alignments)
45.238 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGFYYSCTLYDANGNP.....TIVNGPYGLRHSRLSTVFT 300

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq.36:*

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 1630 | 100.0 | 490 | 21 | Y54122 | Amino acid sequen |
| 2 | 1624 | 99.6 | 476 | 21 | Y54123 | A mannannase-link |
| 3 | 1283 | 78.7 | 468 | 21 | Y44496 | Bacillus agaradher |
| 4 | 1283 | 78.7 | 468 | 21 | Y54125 | Amino acid sequen |
| 5 | 1283 | 78.7 | 493 | 21 | Y44495 | Bacillus agaradher |
| 6 | 1283 | 78.7 | 493 | 21 | Y54124 | Amino acid sequen |
| 7 | 1136 | 69.7 | 331 | 21 | Y54127 | Amino acid sequen |
| 8 | 1075 | 66.0 | 369 | 21 | Y54129 | Amino acid sequen |
| 9 | 937.5 | 57.5 | 320 | 21 | Y54132 | Amino acid sequen |
| 10 | 522 | 32.0 | 188 | 21 | Y54133 | Amino acid sequen |
| 11 | 369 | 22.6 | 305 | 21 | Y54130 | Amino acid sequen |
| 12 | 318 | 19.5 | 132 | 21 | Y54131 | Amino acid sequen |

| | | | | | |
|----|-------|------|------|----|--------|
| 13 | 202 | 12.4 | 410 | 18 | W12378 |
| 14 | 202 | 12.4 | 411 | 18 | W12379 |
| 15 | 202 | 12.4 | 411 | 18 | W12381 |
| 16 | 202 | 12.4 | 412 | 18 | W12380 |
| 17 | 201 | 12.3 | 409 | 18 | W23601 |
| 18 | 201 | 12.3 | 400 | 18 | W22521 |
| 19 | 201 | 12.3 | 400 | 19 | W57431 |
| 20 | 201 | 12.3 | 462 | 19 | W57433 |
| 21 | 200 | 12.3 | 409 | 9 | P81943 |
| 22 | 191 | 11.7 | 467 | 17 | W05731 |
| 23 | 191 | 11.7 | 467 | 17 | W00382 |
| 24 | 174.5 | 10.7 | 499 | 14 | R42122 |
| 25 | 171 | 10.5 | 352 | 20 | Y08472 |
| 26 | 156 | 9.6 | 1010 | 19 | W34989 |
| 27 | 152 | 9.3 | 800 | 8 | P70420 |
| 28 | 152 | 9.3 | 822 | 13 | R26021 |
| 29 | 149.5 | 9.2 | 357 | 16 | R77394 |
| 30 | 149.5 | 9.2 | 941 | 11 | R07478 |
| 31 | 149.5 | 9.2 | 941 | 16 | R77395 |
| 32 | 142 | 8.7 | 358 | 19 | W39262 |
| 33 | 142 | 8.7 | 521 | 17 | R89927 |
| 34 | 142 | 8.7 | 562 | 21 | Y69508 |
| 35 | 135 | 8.3 | 551 | 18 | W18790 |
| 36 | 127 | 7.8 | 472 | 19 | W37243 |
| 37 | 126 | 7.7 | 302 | 19 | W43909 |
| 38 | 124 | 7.6 | 484 | 19 | W43910 |
| 39 | 118.5 | 7.3 | 956 | 19 | W48874 |
| 40 | 118 | 7.2 | 476 | 19 | W37241 |
| 41 | 112.5 | 6.9 | 360 | 19 | W49870 |
| 42 | 112.5 | 6.9 | 360 | 19 | W37242 |
| 43 | 111.5 | 6.8 | 360 | 19 | W34566 |
| 44 | 110.5 | 6.8 | 532 | 12 | R13229 |
| 45 | 110 | 6.7 | 531 | 16 | W01503 |

ALIGNMENTS

| | |
|--------|--------------------------------------------------------------------------|
| RESULT | 1 |
| Y54122 | Y54122 standard; Protein; 490 AA. |
| XX | XX |
| AC | Y54122; |
| XX | XX |
| DE | 27-MAR-2000 (first entry) |
| XX | XX |
| XX | Amino acid sequence of a Bacillus mannanase enzyme. |
| KW | Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase; |
| KW | endo-1,4-mannanase; Bacillus sp. 1633; galactomannan; glucomannan; |
| KW | 1,4-beta-D-mannosidic linkage; mannan; galactomannan; yam; fabricic; |
| KW | galactoglucomanan; cellulosic fibre; synthetic fibre; yarn; fabricic; |
| KW | printing paste; plant material degradation; recycled waste paper; |
| KW | paper making pulp; guar; locust bean gum; thickener; viscosity; |
| KW | mannan-containing food; coffee extract; cleaning composition; |
| KW | machine washing; hard-surface cleaner; dishwashing; oral; dental; |
| KW | contact lens; body-care composition; fabric softener; oil well drilling; |
| KW | subterranean formation fracture. |
| XX | XX |
| OS | Bacillus sp. |
| XX | XX |
| FH | Key |
| FT | Peptide |
| FT | /note- "signal peptide" |
| FT | Domain |
| FT | /note- "catalytic domain" |
| FT | Misc-difference |
| FT | /note- "encoded by CTT" |
| FT | Domain |
| FT | /note- "linker" |
| FT | Domain |
| FT | /note- "domain of unknown function" |
| XX | XX |

P300-Ce1B fusion c
P300-Ce1B fusion c
P300-Ce1B fusion c
P300-Ce1B fusion c
Bacillus agaradher
Bacillus agaradher
Bacillus agaradher
Cloned alkaline en
Sequence of alkali
Cellulase. Bacill
Bacillus cellulase
NK-1 cellulase. B
Actinomycete sp. 3
Teredinbacter end
Sequence encoded b
Alkaline cellulase
Fragment of alkali
Cellulase. Bacill
Full length Bacill
A. cellulolyticus
A. cellulolyticus
Acidothermus cellu
Corrected Bacillus
Globoidea rostrchl
Heterodera glycine
Heterodera glycine
Bankia gouldi glyc
Heterodera glycine
Thermotoga OCI/4V
Globoidea rostrchl
Thermotoga OCI/4V
Endoglucanase enco
60 kd endoglucanas

| Query Match | Best Local Similarity | 100.0% | Score 1630: | DB 21: | Length 490: |
|--------------|-----------------------------------------------------------------|--------|-------------|--------|----------------|
| Matches 300: | Conservative | 0: | Mismatches | 0: | Indels 0: Gaps |
| 1 | NANSGFYVSGTLLYDANGNPFVNRGINHGHWYKDDATTAIEGIANTGANTVRIYVSDG | 60 | | | |
| 31 | nanstgyvsgtlllydangnpgfvmrglnghawykdgdaatlaiegiantgantvriyvsdgs | 90 | | | |
| 61 | QMTKDOIHYRANLISLAEDNHLVAVPEVHATGTYDSIASINRANVDWIERSALLIKERT | 120 | | | |
| 91 | qwtkdoidhyranlislaidnhlvavpevhatacydsiasinrvdwiemrsallikert | 150 | | | |
| 121 | VIININENFWSGEGDAMADGYKQALPRLNAGINHTIAMDAAGWQGFPOSIDYRENFY | 180 | | | |
| 151 | viininenfwsgegdamadgykqalprlnaginhntiamdaagwqgfposidyrenfy | 210 | | | |
| 181 | MADPQRMTFSIMHYEYAGNMSQVTRTNDIVLNOQLALYIGEGHRTHTGVDYDEATIMS | 240 | | | |

[illegible]

CC synthetic fibres, yarn or (non)woven fabrics removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulp
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanaases
CC are also used to process coffee extracts (to inhibit gel formation);
CC in cleaning compositions (for machine washing of fabrics, as
CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
CC dental, contact lens or body-care compositions) where they remove
CC mannan-containing soils and prevent binding of some soils to
CC celluloses; and in fabric softeners. They can also be used in oil
CC well drilling to fracture subterranean formations.

SQ Sequence 476 AA;

| | | | | | | | |
|-----------------------|--------|--------------|----------|------------|----|--------|-----|
| Query Match | 99.6% | Score | 1624 | DB | 21 | Length | 476 |
| Best Local Similarity | 100.0% | Pred. No. | 2.6e-140 | | | | |
| Matches | 299 | Conservative | 0 | Mismatches | 0 | Indels | 0 |
| | | | | | | Gaps | 0 |

| | | | |
|----|-----|-----------------------------------------------------------------|-----|
| QY | 2 | ANSQGVYSGTLLYDANGNPFMRGIRINGHAMVYKQAOATAEGJLANTGAMVRYLYLSPGGO | 61 |
| Db | 1 | ansgfyrsgtcllydangnprfmrjimginhaykqkqatlaegjlanfjancvrylyldsgqg | 60 |
| QY | 62 | WTKDDIHTVRNLISLAEDNHLVAAPVEVDATGYDSIASLNRANDYWTMRMSALIGKEDY | 121 |
| Db | 61 | wtkddihvrrllslsleednhlvavpvhdatqyslaslravdywtmrmsallgkedv | 120 |
| QY | 122 | IINIANEMFSGNEDDANADQYKQALPRLRNAGLWHTLMDNADMGORPOSTIHQYGRVEN | 181 |
| Db | 121 | iiianemfsgweddaedqykqalprlrnaglnhtlmwdaagsgqgppsihnygrvtn | 180 |
| QY | 182 | ADPORNTWFSIHMEYVGNAGNSQVRYRIDRYLNDOLALVYGEFHRTINDVDVETATISY | 241 |
| Db | 181 | adpqrntwfsihmyeyvgngasqyrtidrylndqalavlygefhrntngdvdeatimsy | 240 |
| QY | 242 | SEQSGVGLAMSKGNBPWEYETDLSMDMAGNNTLTAAGNTIVNGPYLSLRRTSRLSYTYT | 300 |
| Db | 241 | seqsgvglamsvngnpweyeldlsmdwagnnltlaagntivngpylslrstsrlstytc | 299 |

| | |
|--------|-----------------------------------|
| RESULT | 3 |
| Y44496 | |
| ID | Y44496 standard; Protein; 468 AA. |

AC Y44496;

DT 27-MAR-2000 (first entry)

DE *Bacillus agaradherens* Clone MB594, Mannanase enzyme

KM Mannanase; endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
KW detergent composition; mid-branched anionic surfactant; washing

05 *Bacillus agaradherens*.
05 Synthetic.

| Key | Location/Qualifiers |
|------------------|---------------------|
| Peptide | 1..31 |
| Signal_peptide | /label= |
| Protein | 32..468 |
| Mature_Mannanase | /label= |

PN W09964552-A1.

PD 16-DEC-1999.

PF 10-JUN-1998; 98WO-US12026.

PR 10-JUN-1998; 98WO-US12026.

PA (PROC) PROCTER & GAMBLE CO.

XX Bettiol JP, Thoen CAJK,
PI

DR WPI; 2000-116536/10.

DR N-PSDB; Z29846

PT Detergent composition for removing greasy stains such as cosmetics,
PT food stains and body soils -

PS Disclosure; Page 103-104; 113pp; English

The present sequence is the B agaradhaerens clone M559, alkaline mannanase enzyme. Clone M559 is derived from B. agaradhaerens strain NCIMB 40482 genomic DNA. Mannanase shows maximum activity at pH ranging from 7.5-10.5. It can be used in a detergent composition along with a mild-branded anionic surfactant. The detergent composition may be used for washing purposes, soaking/pre-treatment of stained fabric, hard surface cleaning and for removal of cosmetic and/or food stains. This composition provides excellent cleaning effect at low temperature.

SQ Sequence 468 AA

| | | | | |
|-----------------------|-----------------|--------------------|----------|------------|
| Query Match | 78.7% | Score 1283 | DB 21 | Length 468 |
| Best Local Similarity | 77.3% | Pred. No. 3.5e+109 | | |
| Matches 231 | Conservative 30 | Mismatches 38 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|------------------------------------------------------------------------------------------------------|-----|
| QY | 2 | ANPEYVSGTLYANGNPPMRBINGHAMUKOATATEGIANCANVRIYLSGQ | 61 |
| | | ::: | |
| Db | 32 | asctgydggnllydangprrvmyclmghnauykdstacta pra eegqantlrvly sdcgq | 91 |
| QY | 62 | WTKODIHTVRNLI SLAEDNHL VAPEYHADATGYDS SLNKRAVDY WKMSALIGKEDTV | 121 |
| | | | |
| Db | 92 | wekoddltlrfv elaeqnmkwa vver hd atgr dsr dl n rvayv emkda l gk edtv | 151 |
| QY | 122 | I I N A N E F T G W E D N A D G K O A I P R N A N G H T M D A G G Q P O S H O V G E V N | 181 |
| | | | |
| Db | 152 | l l n a n e y g w d s a w d g y d v l p r t a g l c h t m w d a a g y q p s h d y g d v n | 211 |
| QY | 182 | ADPO RN T F S I H H E Y A G N S Q V R T N D R N O D L V I G E F H R T N G D V E A T I T S Y | 241 |
| | | | |
| Db | 212 | a d r l k h m f s h m y e y g d a n t r v n d r v l d q d a l v g e f h r n h t d g d e d t l s y | 271 |
| QY | 242 | SE Q G V G H A N S K G N C P E W E T L D S D N A G N T I T A N G I T V N C P G L R E T S R L S T Y T | 300 |
| | | | |
| Db | 272 | s e e g t y a w s v g n s t e w y d l s c d e a g h l t d g n l v h a d l g e s k s p v t | 330 |

| | | |
|--------|--------|---------------------------|
| RESULT | 4 | , |
| Y54125 | | , |
| ID | Y54125 | standard; Protein; 468 AA |

AC Y541257

DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a *Bacillus mannanase* enzyme

KW Mannanase: mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW gear; locust bean gum; mannan containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.xx

05 Synthetic.

OS *Bacillus agaradhaerens*.

PN WO9964619-A2

16-DEC-1999.
 XX PF 10-JUN-1999; 99WO-DK00314.
 XX PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 09-MAR-1999; 99DK-0000309.
 PR 10-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 11-MAR-1999; 99US-0123641.
 XX PA (NOVO) NOVO-NORDISK AS.
 XX PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX DR WPI: 2000-105891/09.
 XX DR N-PSDB; 245338.
 PT New mannases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX PS Example 5; Page 215-216; 242pp; English.
 XX The present sequence represents a mannanase (also known as mannan
 CC endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The present mannanase is a synthetic variant of the mannanase of
 CC T34124, in which the C-terminus of the protein was changed due to
 CC design of a lower PCR primer used for amplification. The mannanase
 CC hydrolyses galactomannans. Specifically, mannases hydrolyse
 CC 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation);
 CC in cleaning compositions (for machine washing of fabrics, as
 CC hard-surface cleaners, for hand or machine dishwashing, also in oral,
 CC dental, contact lens or body-care compositions) where they remove
 CC mannan-containing soils and prevent binding of some soils to
 CC cellulosics; and in fabric softeners. They can also be used in oil
 CC well drilling to fracture subterranean formations.
 XX SQ Sequence 468 AA;

Query Match 78.7%; Score 1283; DB 21; Length 468;
 Best Local Similarity 77.3%; Pred. No. 3.5e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

2 ANSGFYVSGTLLYDANGNPFVARGINHGHWYKDOATTAIEGIANTGANTVRIYLSDGQ 61
 Db 32 astgfyvgnltlydangpffvmrginhghaykdstastlpaaleagantlrlvlsdgq 91
 Qy 62 WTKDITHTVRLNLSIAEDNHLVAVPEVHDATGYSIASLNRAVDYWIEMRSALIGKEDTV 121
 Db 92 wkddidtlrtvrlaegnkhwavvevhdatsrdsdlnravdywlemkdaligkdedtv 151
 Qy 122 IININEMWGSWGDWAGYKQAIPTRLNAGINHTLMDAGNGQFPOSIDYGREVN 181
 Db 152 IININEMWGSWGDWAGYKQAIPTRLNAGINHTLMDAGNGQFPOSIDYGREVN 211
 Qy 182 ADPORTNFSIMHYEYAGNASOVRTNIDRVINODLATVIGFGRHRTGVDDEATINSY 241

212 adpikntfslmhyegagdantrvsnidrvldqlalvigeifgnrthdgvdcdtlisly 271
 Qy 242 SEORGVGWLAWSMKNGNPEWEXYDLSNDMAGNNLTAMGNTIVNGPYGRSRLSTVFT 300
 Db 272 seetgtglawswkngnstewdyldlsedwagqhltdwgnrlvhgadqigetskptvft 330

RESULT 5
 Y44495
 ID Y44495 standard; Protein; 493 AA.
 AC Y44495;
 DT 27-MAR-2000 (first entry)
 DE Bacillus agaradherens NCIMB 40482, Mannanase enzyme.
 KW Mannanase: endo-1,4-beta-mannosidase; E.C. 3.2.1.78; cleaning;
 KW detergent composition; mid-branched anionic surfactant; washing;
 KW cosmetic stain; food stain.
 OS Bacillus agaradherens.
 FH Key Location/Qualifiers
 FT Peptide 1..32
 FT /label= Signal_peptide
 FT 33..343
 FT Protein /label= Mature_Mannanase
 XX PN W09964552-A1.
 XX PD 16-DEC-1999.
 XX PR 10-JUN-1998; 98WO-US12026.
 XX PR 10-JUN-1998; 98WO-US12026.
 XX PR (PROC) PROCTER & GAMBLE CO.
 XX PI Bettiol JP, Thoen CAUK;
 XX DR WPI: 2000-116536/10.
 XX DR N-PSDB; 229845.
 PT Detergent composition for removing greasy stains such as cosmetics,
 PT food stains and body soils
 XX PS Disclosure; Page 102; 113pp; English.
 CC The present sequence is the B. agaradherens NCIMB 40482, alkaline
 CC mannanase enzyme. Mannanase shows maximum activity at pH ranging
 CC from 7.5-10.5. It can be used in a detergent composition along with a
 CC mid-branched anionic surfactant. The detergent composition may be used
 CC for washing purposes, soaking/pre-treatment of stained fabric, hard
 CC surface cleaning and for removal of cosmetic and/or food stains. This
 CC composition provides excellent cleaning effect at low temperature.
 XX SQ Sequence 493 AA;

Query Match 78.7%; Score 1283; DB 21; Length 493;
 Best Local Similarity 77.3%; Pred. No. 3.7e-109;
 Matches 231; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

2 ANSGFYVSGTLLYDANGNPFVARGINHGHWYKDOATTAIEGIANTGANTVRIYLSDGQ 61
 Db 32 astgfyvgnltlydangpffvmrginhghaykdstastlpaaleagantlrlvlsdgq 91
 Qy 62 WTKDITHTVRLNLSIAEDNHLVAVPEVHDATGYSIASLNRAVDYWIEMRSALIGKEDTV 121
 Db 92 wkddidtlrtvrlaegnkhwavvevhdatsrdsdlnravdywlemkdaligkdedtv 151

| | |
|---------------------------|-----------------------------------------------------------------------------|
| PS | Claim 35; Page 213-214; 242pp; English. |
| XX | |
| CC | The present sequence represents a mannanase (also known as mannan |
| CC | endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The |
| CC | mannanase hydrolyses galactomannans. Specifically, mannases hydrolyse |
| CC | 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, |
| CC | and galactoglucomannans. The mannanase protein, or preparations |
| CC | containing it, are used to improve properties of cellululosic or |
| CC | synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based |
| CC | sizes or printing pastes). They are also used to degrade or modify |
| CC | plant materials (particularly recycled waste paper, paper making pulp, |
| CC | or material containing guar or locust bean gums (thickeners), or to |
| CC | reduce viscosity of mannan-containing foods or feeds). The mannases |
| CC | are also used to process coffee extracts (to inhibit gel formation); |
| CC | in cleaning compositions (for hand or machine washing of fabrics, as |
| CC | hard-surface cleaners, for food or machine dishwashing, also in oral, |
| CC | dental, contact lens or body-care compositions) where they remove |
| CC | mannan-containing soils and prevent binding of some soils to |
| CC | cellulosics) and in fabric softeners. They can also be used in oil |
| CC | well drilling to fracture subterranean formations. |
| CC | |
| SQ | Sequence 493 AA; |
| Query Match | 78.7%; Score 1283; DB 21; Length 493; |
| Best Local Similarity | 77.3%; Pred. No. 3.7e-109; |
| Matches 231; Conservative | 30; Mismatches 38; Indels 0; Gaps |
| QY | 2 ANSEFFVSGTTVDANGNPVMGIVNGHAWMYDQAQTALIEGANTGANTRVIVSDGGQ 61 |
| Db | 32 astgfydgntllldangqprfvmglbhghawdydstalpaaleegantlrivtsdgg 91 |
| QY | 62 WTKODITFVNRLSLAEDNHLVAPEVGHATGDSIASINRAVDYWEKRSALLKEEDTV 121 |
| Db | 92 wekdittitrevleleeqnkmvavevhdtgtgdsrdlnravdyvlenkddlkgedtv 151 |
| QY | 122 IINTANEMFGSWEGDAWADGYKQALPRLNAGLNIHTLMYDAAGWGQFPOSIDHYGREVFN 181 |
| Db | 152 llnlanewygswdgsawadgyldtvlpkrlrdaqlthtlmwaagwgvgypslndygqvfhn 211 |
| QY | 182 ADPORNTMFSTHWTEYGVGNASOVRNRINDRYLNMQDALVIYGEGRHRTGTGDVENTIMSY 241 |
| Db | 212 adplknkfmsflmyeyagsgdancvtcrsnldrvldqdialylgetghrhrtgdvdcdtllsy 271 |
| QY | 242 SEQGVCGLMWSMGNGPEWEYLDLSNDMAGNULZMWGTYINGPYGLRETSRLSTVEFT 300 |
| Db | 272 seeqtgvlewskwnstewdyldlsedwaqghltowgnirinhgaadylgetskptvftl 330 |
| RESULT | 7 |
| ID | YS4127 |
| XX | YS4127 standard; Protein: 331 AA. |
| AC | YS4127; |
| DT | 27-MAR-2000 (first entry) |
| DE | Amino acid sequenc of a Bacillus mannanase enzyme. |
| KX | Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase; |
| KW | endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage; |
| KW | mannan; galactomannan; glucomannan; galactoglucomannan; cellululosic fibre; |
| KW | synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; |
| KW | plant material degradation; recycled waste paper; paper making pulp; |
| KW | guar; locust bean gum; mannan-containing food; coffee extract; |
| KW | cleaning composition; machine washing; hard-surface cleaner; |
| KW | dishwashing; oral; dental; contact lens; body-care composition; |
| KW | fabric softener; oil well drilling; subterranean formation fracture. |
| OS | Bacillus sp. |
| PH | Key Location/Qualifiers |

```

FT Peptide 1..32
FT /note= "signal peptide"
FT Domain 33..331
FT /note= "catalytic domain"
PN MO9964619-A2.
XX 16-DEC-1999.
PD
XX
PF 10-JUN-1999; 99WO-DK00314.
XX
PR 10-JUN-1998; 98US-0111256.
PR 20-OCT-1998; 98DK-0001340.
PR 20-OCT-1998; 98DK-0001341.
PR 28-OCT-1998; 98US-0105970.
PR 28-OCT-1998; 98US-0106054.
PR 23-DEC-1998; 98DK-0001725.
PR 05-MAR-1999; 99DK-0000306.
PR 05-MAR-1999; 99DK-0000307.
PR 05-MAR-1999; 99DK-0000308.
PR 09-MAR-1999; 99US-0123543.
PR 10-MAR-1999; 99US-0123623.
PR 10-MAR-1999; 99US-0123641.
PR 11-MAR-1999; 99US-0123642.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Kaupinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME,
XX WPI: 2000-105891/09.
XX N-PSDB; Z45340.
XX
XX New mannanses for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions
XX
XX Claim 35; Page 220-221; 242pp; English.
XX
XX The present sequence represents a Bacillus mannanase (also known as
XX mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
XX The mannanase hydrolyses galactomannans. Specifically, mannanases
XX hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
XX glucomannans, and galactoglucomannans. The mannanase protein, or
XX preparations containing it, are used to improve properties of cellulosic
XX or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
XX sizes or printing pastes). They are also used to degrade or modify
XX plant materials (particularly recycled waste paper, paper making pulps,
XX or material containing guar or locust bean gums (thickeners), or to
XX reduce viscosity of mannan-containing foods or feeds). The mannanases
XX are also used to process coffee extracts (to inhibit gel formation); in
XX cleaning compositions (for machine washing of fabrics, as hard-surface
XX cleaners, for hand or machine dishwashing, also in oral, dental, contact
XX lens or body-care compositions) where they remove mannan-containing
XX soils and prevent binding of some soils to cellulosics; and in fabric
XX softeners. They can also be used in oil well drilling to fracture
XX subterranean formations.
XX
XX Sequence 331 AA:
SQ

```

```

DB 154 ininewyagwddgavargynalrqlrlnaglstftfnwdaagayqypgsvvdygve1na 213
OY 183 DPQRNMFSLHMYEYAGGNASQVFTNIDRYLNDLALVIGEPGHRHNGVDENTINSYS 242
DB 214 dpgtrntmfvmyeyagddentvrrnidsllsqnlavlgfghwbydgvdedtl1s 273
OY 243 EQRGVGLAWMSKNGKPEWEYLDLSNDMAGNLTAMGNTIVNGPYGLRSTSTVE 299
DB 274 qqrnvgyawswagsegveyldlsndfagurltwgdrlvngpnlrftskssvf 330

RESULT 8
Y54129
ID Y54129 standard; Protein; 369 AA.
XX
XX Y54129;
AC
XX
DT 27-MAR-2000 (first entry)
XX
DE Amino acid sequence of a Bacillus sp. AA349 mannanase enzyme.
XX
XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
XX endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
XX mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
XX synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
XX guar; locust bean gum; mannan-containing food; coffee extract;
XX cleaning composition; machine washing; hard-surface cleaner;
XX dishwashing; oral; dental; contact lens; body-care composition;
XX fabric softener; oil well drilling; subterranean formation fracture.
XX
XX Bacillus sp.
OS
XX
XX Key location/Qualifiers
XX Peptide 1..25
XX /note= "signal peptide"
XX
XX Domain 26..67
XX
XX Domain 68..369
XX /note= "N-terminal domain of unknown function"
XX /note= "catalytic domain"
PN MO9964619-A2.
XX
XX 16-DEC-1999.
PD
XX
XX 10-JUN-1999; 99WO-DK00314.
XX
XX 10-JUN-1998; 98US-0111256.
XX 20-OCT-1998; 98DK-0001340.
XX 20-OCT-1998; 98DK-0001341.
XX 28-OCT-1998; 98US-0105970.
XX 28-OCT-1998; 98US-0106054.
XX 23-DEC-1998; 98DK-0001725.
XX 05-MAR-1999; 99DK-0000306.
XX 05-MAR-1999; 99DK-0000307.
XX 05-MAR-1999; 99DK-0000308.
XX 05-MAR-1999; 99DK-0000309.
XX 09-MAR-1999; 99US-0123543.
XX 10-MAR-1999; 99US-0123623.
XX 10-MAR-1999; 99US-0123641.
XX 11-MAR-1999; 99US-0123642.
XX
XX (NOVO ) NOVO-NORDISK AS.
XX
XX Kaupinen MS, Schnelein M, Schnorr K, Andersen LN, Bjornvad ME;
XX WPI: 2000-105891/09.
XX N-PSDB; Z45342.
XX
XX New mannanses for treatment of textiles, plant material and coffee
XX extract, and in cleaning compositions
XX

```

PS Claim 35; Page 224-225; 242pp; English.

XX The present sequence represents a mannanase enzyme (also known as
CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
CC The mannanase hydrolyses galactomannans. Specifically, mannanases
CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
CC glucomannans, and galactoglucomannans. The mannanase protein, or
CC preparations containing it, are used to improve properties of cellulosic
CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
CC sizes or printing pastes). They are also used to degrade or modify
CC plant materials (particularly recycled waste paper, paper making pulps,
CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods or feeds). The mannanases
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.

XX Sequence 369 AA;

Query Match 66.0%; Score 1075; DB 21; Length 369;
Best Local Similarity 66.2%; Pred. No. 2.5e-90;
Matches 196; Conservative 37; Mismatches 63; Indels 0; Gaps 0;

QY 4 SGFVSGTLLDANGNPFVNGINGHGMAYKDQATTAEGTANTGANTVRIVLSDGQGT 63
DB 72 ngfyldgtitlydangkprmgjnhghwkyphletamealadganslrvlvdgqwt 131
QY 64 KDIHTVRLNISLADNHLVAVPEVHDATGYDSIASLNRADVIMEMKSLIGEDTVII 123
DB 132 kddvdevaklisladekhsvalaelehdldgldlepllktydyfelkdaigkcdkvi 191
QY 124 NIANEWGSWEGDAMAGCYKRAIPRLRAGNHTLMDAAGWGPPOSIHYGAEVFNAD 183
DB 192 nlsmewgswwsegwadgykkaiprlreaglkhlmwdaagwqgiprsihkglevtsnd 251
QY 184 PORNTWFSIHMYEXAGNASQVFTNIDVLTMODLALVIGEGFHHHTNGDVDEATIMSTVE 243
DB 252 plktmteihmyewagunpvgkndlgvleknalavlygeghbhhygdvavdtllshse 311
QY 244 QRGVGTAWSKRGNGPEWEYLDISNDNAGNNLTAMGNTIVNGPYGLRSTSLSTVE 299
DB 312 kydvglawswghnsgvgyldltdfsgtqltewgerivhnpuglketseivsy 367

RESULT 9
Y54132
ID Y54132 standard; Protein; 320 AA.

AC Y54132;

DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a Bacillus sp. mannanase enzyme.

XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KM plant material degradation; recycled waste paper; paper making pulp;
KM guar; locust bean gum; mannan-containing food; coffee extract;
KM cleaning composition; machine washing; hard-surface cleaner;
KM dishwashing; oral; dental; contact lens; body-care composition;
KM fabric softener; oil well drilling; subterranean formation fracture.

XX Bacillus sp.

XX Key Location/Qualifiers
FH 1..28
FT Peptide

FT /note- "signal peptide"
FT 29..320
FT /note- "catalytic domain"

PN WC0964619-A2.

PD 16-DEC-1999.

PF 10-JUN-1999; 99WO-DK00314.

PR 10-JUN-1998; 98US-011256.

PR 20-OCT-1998; 98DK-0001340.

PR 20-OCT-1998; 98DK-0001341.

PR 28-OCT-1998; 98US-0105970.

PR 28-OCT-1998; 98US-0106054.

PR 23-DEC-1998; 98DK-0001725.

PR 05-MAR-1999; 99DK-0000306.

PR 05-MAR-1999; 99DK-0000307.

PR 05-MAR-1999; 99DK-0000308.

PR 05-MAR-1999; 99DK-0000309.

PR 09-MAR-1999; 99US-0123543.

PR 10-MAR-1999; 99US-0123623.

PR 10-MAR-1999; 99US-0123641.

PR 11-MAR-1999; 99US-0123642.

PA (NOVO) NOVO-NORDISK AS.

PI Kaupinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;

XX WPI: 2000-105891/09.

DR N-PSDB; 245345.

PT New mannanases for treatment of textiles, plant material and coffee

XX extract, and in cleaning compositions

PS Claim 35; Page 228-229; 242pp; English.

XX The present sequence represents a mannanase enzyme (also known as

CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).

CC The mannanase hydrolyses galactomannans. Specifically, mannanases

CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,

CC glucomannans, and galactoglucomannans. The mannanase protein, or

CC preparations containing it, are used to improve properties of cellulosic

CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

CC sizes or printing pastes). They are also used to degrade or modify

CC plant materials (particularly recycled waste paper, paper making pulps,

CC or material containing guar or locust bean gums (thickeners), or to

CC reduce viscosity of mannan-containing foods or feeds). The mannanases

CC are also used to process coffee extracts (to inhibit gel formation); in

CC cleaning compositions (for machine washing of fabrics, as hard-surface

CC cleaners, for hand or machine dishwashing, also in oral, dental, contact

CC lens or body-care compositions) where they remove mannan-containing

CC soils and prevent binding of some soils to celluloses; and in fabric

CC softeners. They can also be used in oil well drilling to fracture

CC subterranean formations.

XX Sequence 320 AA;

Query Match 57.5%; Score 937.5; DB 21; Length 320;
Best Local Similarity 59.7%; Pred. No. 7.2e-78;
Matches 176; Conservative 39; Mismatches 79; Indels 1; Gaps 1;

QY 1 NANGFVSGTLLDANGNPFVNGINGHGMAYKDQATTAEGTANTGANTVRIVLSDG 60

DB 27 sagdfvntkgtellidkngddpymgvnhghswfvgdleealpaaegantvrvlvsng 86

QY 61 QWTKDHTVRLNISLADNHLVAVPEVHDATGYDSIASLNRADVIMEMKSLIGEDT 120

DB 87 gwexddaselarvlatetgylttvlewhdatsgndpddldkavdylemadvlkgtedr 146

QY 121 VIINIANEMFGSWEGDAMAGCYKRAIPRLRAGNHTLMDAAGWGPPOSIHYGAEV 180

Db 147 vlnlanewygvawrsdwaesayagaprlrzsaglahcllvdaagwgyqypasibhergadvf 206
 QY 181 NADQORNTMSEIHMYEYAGNAGNAGVFNIDVNLNODALVYGEHRTNGDVDEATIMS 240
 Db 207 asqdkntmrtshmyeyagadravsenldgvleenaivgeghhghdgvdedadlla 266
 QY 241 YSEQRGVGLWMSKNGPEWEYIDLSDNMGNNLTANGNTIVNGPYGLRETSRL 295
 Db 267 ytaerqvgwlawswygnaggyeyldltegpag-pltswgerliyvgemglkvldhl 320

RESULT 10
 Y54133
 ID Y54133 standard; Protein; 188 AA.
 XX
 AC Y54133;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Amino acid sequence of a partial *Bacillus* sp. mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KM plant material degradation; recycled waste paper; paper making pulp;
 KM guar; locust bean gum; mannan-containing food; coffee extract;
 KM cleaning composition; machine washing; hard-surface cleaner;
 KM dishwashing; oral; dental; contact lens; body-care composition;
 KM fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS *Bacillus* sp.
 XX
 PN W09964619-A2.
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 98US-0123543.
 PR 10-MAR-1999; 98US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauplien MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI, 2000-105891/09.
 DR N-PSDB; 245346.
 XX
 PT New mannases for treatment of textiles, plant material and coffee
 XX extract, and in cleaning compositions
 PS
 PS Claim 35; Page 229-230; 242pp; English.
 CC The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing; also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 XX
 SQ Sequence 188 AA;
 XX

Query Match 32.0%; Score 522; DB 21; Length 188;
 Best Local Similarity 59.9%; Pred. No. 3e-40;
 Matches 97; Conservative 20; Mismatches 45; Indels 0; Gaps 0;

QY 1 NANGFYVSGTLYDANGNPFVARGIHGHAWYKQDQATTBEGANMGATVRYVSDGG 60
 Db 27 saqgfiwkgfclldkngdpyvmrghshwfkqdllealpaiaeagantvrlvsng 86
 QY 61 QWTKDDIHTVKNLSTLAEDNHLVAVPEVHDATGYDSIASINRAVDYWIENRSALIGKEDT 120
 Db 87 qwekdaseelarvvaateyglttvlevhdatsdnpdddkavdytemadvlkgredr 146
 QY 121 VIINANWFGSWEGDAMADGKQAIPLRLNAGINHTLVADA 162
 Db 147 vlnlanewygvawrsdwaesayagaprlrzsaglahcllvda 188

RESULT 11
 Y54130
 ID Y54130 standard; Protein; 305 AA.
 XX
 AC Y54130;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Amino acid sequence of a partial *Bacillus* sp. mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KM endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KM mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KM synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KM plant material degradation; recycled waste paper; paper making pulp;
 KM guar; locust bean gum; mannan-containing food; coffee extract;
 KM cleaning composition; machine washing; hard-surface cleaner;
 KM dishwashing; oral; dental; contact lens; body-care composition;
 KM fabric softener; oil well drilling; subterranean formation fracture.
 XX
 OS *Bacillus* sp.
 XX
 PN W09964619-A2.
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 98US-0123543.
 PR 10-MAR-1999; 98US-0123623.
 PR 10-MAR-1999; 99US-0123641.

PR 11-MAR-1999; 99US-0123642.
 XX (NOVO) NOVO-NORDISK AS.
 PA
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 DR WPI: 2000-105891/09.
 N-PSDB: 245343.
 XX
 XX
 PT New mannanases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX
 XX
 PS Disclosure: Page 225-226; 242pp; English.
 XX
 XX The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulotics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 XX
 XX Sequence 305 AA;

Query Match 22.6%; Score 369; DB 21; Length 305;
 Best Local Similarity 45.3%; Pred. No. 5,4e-26;
 Matches 73; Conservative 32; Mismatches 46; Indels 10; Gaps 4;

OY 146 IRLNAGLNTMTVDAGQFPOSINHDYGRFVNADPQRTWTFSIHYETAGNANOV 205
 DB 1 IRLNAGLNTMTVDAGQFPOSINHDYGRFVNADPQRTWTFSIHYETAGNANOV 205
 OY 206 RKNIDRVNODIALYIGFGRHRTNCD-----VDEATIMSTSEORGVMANWNGNCP 259
 DB 60 GELGALKGLIYAMIGETGYNGNMLGSGVNAQGAQKGIYMPWSTGND 119
 OY 260 EWEYDL-SNDAGNLTMTWGTIVNGPYGLRSTSLSTVF 299
 DB 120 ANSWIDMTNDW--GLTSWGNLVVNGTNGIRATSPATVF 158

RESULT 12
 Y54131
 ID Y54131 standard; Protein; 132 AA.

AC Y54131;
 DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a partial Bacillus mannanase enzyme.
 XX
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; galactoglucomannan; specifically, mannanases
 KW hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 KW glucomannans, and galactoglucomannans. The mannanase protein, or
 KW preparations containing it, are used to improve properties of cellulosic
 KW or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 KW sizes or printing pastes). They are also used to degrade or modify
 KW plant materials (particularly recycled waste paper, paper making pulps,
 KW or material containing guar or locust bean gums (thickeners), or to
 KW reduce viscosity of mannan-containing foods or feeds). The mannanases
 KW are also used to process coffee extracts (to inhibit gel formation); in
 KW cleaning compositions (for machine washing of fabrics, as hard-surface
 KW cleaners, for hand or machine dishwashing, also in oral, dental, contact
 KW lens or body-care compositions) where they remove mannan-containing
 KW soils and prevent binding of some soils to cellulotics; and in fabric
 KW softeners. They can also be used in oil well drilling to fracture
 KW subterranean formations.
 XX

OS Bacillus clausii.
 XX
 XX W09964619-A2.
 XX
 XX
 PD 16-DEC-1999.
 XX
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 XX 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 XX (NOVO) NOVO-NORDISK AS.
 XX
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX
 XX WPI: 2000-105891/09.
 DR N-PSDB: 245344.
 XX
 XX
 PT New mannanases for treatment of textiles, plant material and coffee
 PT extract, and in cleaning compositions
 XX
 XX
 PS Disclosure: Page 227; 242pp; English.
 XX
 XX

CC The present sequence represents a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulotics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 CC
 XX
 XX Sequence 132 AA;

Query Match 19.5%; Score 318; DB 21; Length 132;
 Best Local Similarity 55.6%; Pred. No. 7,8e-22;
 Matches 60; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

OY 2 ANSGFYSGTILYDANGFNVRGINHGHWYKQDATTATIEGANTGATVRIYVSDGO 61
 DB 23 AGTGFYGTGTLIDDEGPNYVYVNGHGWTKGQIDTALPAATGANTVRIYVNSGQ 82
 OY 62 WKKDIIHYVKNLSLAEDNHLVAYVEVDATGYDSIASLNRAVDWTF 109
 DB 83 WERTDVAEYVLAEEGLAVLVADATGSDPDNLTFAVEYWE 130

RESULT 13
 W12378
 ID W12378 standard; Protein; 410 AA.

```

XX AC W12378;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 1 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. M4 (ATCC 21833).
XX FH Key
XX FT Peptide
XX FT 1..26
XX FT /label= Sig_peptide
XX FT /note= "Hybrid between P300 (aa1-12) and
XX FT CelB (aa13-26) signal peptides"
XX FT Protein
XX FT 27..410
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN W09706181-A1.
XX PD 20-FEB-1997.
XX PE 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENKEL) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB: T63226.
XX PT System for increased expression of cellulase and xylanase in
XX PT Bacillus - contains gene under control of elements from B.
XX PT licheniformis alkaline protease gene
XX PS Disclosure: Fig 7,7A,7B; 37pp; English.
XX CC The polypeptide product (W12378) of P300-CelB fusion construct 1
XX CC (T63226) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. M4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also
XX CC W12379-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 410 AA:
OY Query Match 12.4%; Score 202; DB 18; Length 410;
OY Best Local Similarity 25.0%; Pred. No. 1.4e-10;
OY Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;
DB 41 isngelvdrtgpevqikmshnglqvyg-----qfnyesmklrdd 82
OY 67 -----IHT-----VRNLISIAEDNHLVAVPEVDANGYDSIASLNR 103
DB 83 wgtlvtiraamyssggyljedpsvkkevkeaveaidlglyvlldwhllednphnykeea 142
OY 104 VDWIMRSGALICKEDPTVIININEMFGS---WEGD--AMADGYKQALPRLNAGINHTL 158
DB 143 ktdidem-selygdvnyvlyelanepgsdvtwdnqikpyae---evlpylrndpnll 198

```

```

OY 159 MYDAAGMGQFPQSIDYGRVFNADPQRTKPSIMHYEYAGNASQVRNIDRVNODLA 218
DB 199 lv---gtgtwsgdvh-haadnqltdp--nvmvafhf--yagthgqnldqvayldgga 250
OY 219 LVIGFEGHRTNGD---VDEATI-MSYSEQGVGLWLS 253
DB 251 lfsewgtseatgdggyfideagvldfndernlswans 290
RESULT 14
W12379
ID W12379 standard; Protein; 411 AA.
XX AC W12379;
XX DT 17-JUN-1997 (first entry)
XX DE P300-CelB fusion construct 2 polypeptide product.
XX KM Cellulase; xylanase; alkaline protease; P300; CelB.
XX OS Chimeric Bacillus licheniformis ATCC 53926;
XX OS Chimeric Bacillus sp. M4 (ATCC 21833).
XX FH Key
XX FT Peptide
XX FT 1..26
XX FT /label= Sig_peptide
XX FT /note= "Hybrid between P300 (aa1-5) and
XX FT CelB (aa8-26) signal peptides"
XX FT Protein
XX FT 27..411
XX FT /label= Mat.protein
XX FT /note= "mature CelB cellulase"
XX PN W09706181-A1.
XX PD 20-FEB-1997.
XX PE 08-AUG-1996; 96WO-US12545.
XX PR 07-AUG-1996; 96US-0694346.
XX PR 10-AUG-1995; 95US-0002106.
XX PA (HENKEL) HENKEL CORP.
XX PI Christianson T, Maurer K, Tang MR, Weiss A, Wilson CR;
XX DR WPI: 1997-154208/14.
XX DR N-PSDB: T63227.
XX PT System for increased expression of cellulase and xylanase in
XX PT Bacillus - contains gene under control of elements from B.
XX PT licheniformis alkaline protease gene
XX PS Disclosure: Fig 8,8A,8B; 37pp; English.
XX CC The polypeptide product (W12379) of P300-CelB fusion construct 2
XX CC (T63227) comprises a hybrid signal peptide, formed between the
XX CC signal sequences of Bacillus licheniformis ATCC 53926 (P300)
XX CC alkaline protease and Bacillus sp. M4 CelB cellulase, and the
XX CC mature CelB alkalophilic cellulase. Expression in Bacillus sp.
XX CC host cells of P300-CelB fusion constructs 1-4 (see also W12378,
XX CC W12380-81) provides a 10-40 fold improvement of prodn. of the
XX CC alkalophilic cellulase in comparison to expression of the native
XX CC gene, and an earlier start to the prodn. of enzyme in the
XX CC fermentation process.
XX SQ Sequence 411 AA:
OY Query Match 12.4%; Score 202; DB 18; Length 411;
OY Best Local Similarity 25.0%; Pred. No. 1.4e-10;
OY Matches 70; Conservative 48; Mismatches 98; Indels 64; Gaps 14;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:06:50 ; Search time 154.68 Seconds
(Without alignments)
32.508 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGFYSGTLYDANGNP.....TIVNGPYGLRHSRLSTVFT 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|--------------------|
| 1 | 201 | 12.3 | 400 | US-08-713-298B-2 | Sequence 2, Appl1 |
| 2 | 201 | 12.3 | 400 | US-08-870-180B-2 | Sequence 2, Appl1 |
| 3 | 201 | 12.3 | 400 | US-08-814-052-4 | Sequence 4, Appl1 |
| 4 | 201 | 12.3 | 400 | US-08-812-829-4 | Sequence 4, Appl1 |
| 5 | 201 | 12.3 | 462 | US-08-870-180B-13 | Sequence 13, Appl1 |
| 6 | 191 | 11.7 | 467 | US-08-727-548-2 | Sequence 2, Appl1 |
| 7 | 171.5 | 10.5 | 430 | US-08-924-440-2 | Sequence 2, Appl1 |
| 8 | 142 | 8.7 | 358 | US-08-604-913B-11 | Sequence 11, Appl1 |
| 9 | 142 | 8.7 | 521 | US-08-276-213-3 | Sequence 3, Appl1 |
| 10 | 135 | 8.7 | 551 | US-09-033-537A-1 | Sequence 1, Appl1 |
| 11 | 110 | 6.7 | 531 | US-07-862-588B-7 | Sequence 7, Appl1 |
| 12 | 100 | 6.1 | 429 | US-08-745-977-4 | Sequence 4, Appl1 |
| 13 | 100 | 6.1 | 429 | US-09-040-699A-4 | Sequence 4, Appl1 |
| 14 | 99.5 | 6.1 | 317 | US-09-066-075-2 | Sequence 2, Appl1 |
| 15 | 99.5 | 6.1 | 317 | US-08-581-889-2 | Sequence 2, Appl1 |
| 16 | 99.5 | 6.1 | 317 | US-08-732-433-1 | Sequence 1, Appl1 |
| 17 | 97.5 | 6.0 | 574 | US-08-459-499-12 | Sequence 12, Appl1 |
| 18 | 95.5 | 5.9 | 735 | US-08-313-185-48 | Sequence 48, Appl1 |
| 19 | 95.5 | 5.9 | 735 | US-08-459-499-9 | Sequence 9, Appl1 |
| 20 | 94 | 5.8 | 398 | US-08-853-659A-45 | Sequence 45, Appl1 |
| 21 | 93.5 | 5.7 | 566 | US-07-862-588B-4 | Sequence 4, Appl1 |
| 22 | 90.5 | 5.6 | 735 | US-08-459-499-12 | Sequence 12, Appl1 |
| 23 | 88 | 5.4 | 531 | US-08-531-601-1 | Sequence 1, Appl1 |
| 24 | 88 | 5.4 | 531 | US-08-859-032-1 | Sequence 1, Appl1 |
| 25 | 86 | 5.4 | 535 | US-08-737-597-10 | Sequence 10, Appl1 |
| 26 | 86 | 5.3 | 461 | US-08-672-571A-3 | Sequence 3, Appl1 |
| 27 | 85 | 5.3 | 490 | US-08-672-571A-1 | Sequence 1, Appl1 |
| 28 | 85.5 | 5.2 | 385 | US-08-387-942C-26 | Sequence 26, Appl1 |

| | | | | | |
|----|------|-----|------|-------------------|--------------------|
| 29 | 85.5 | 5.2 | 997 | US-08-387-942C-4 | Sequence 4, Appl1 |
| 30 | 85.5 | 5.2 | 1248 | US-08-348-353-17 | Sequence 17, Appl1 |
| 31 | 85.5 | 5.2 | 1248 | US-08-465-965-17 | Sequence 17, Appl1 |
| 32 | 85.5 | 5.2 | 1248 | US-08-465-966-17 | Sequence 1, Appl1 |
| 33 | 85 | 5.2 | 881 | US-08-333-901-1 | Sequence 1, Appl1 |
| 34 | 85 | 5.2 | 881 | US-08-456-582-1 | Sequence 1, Appl1 |
| 35 | 85 | 5.2 | 881 | US-08-898-789-1 | Sequence 1, Appl1 |
| 36 | 85 | 5.2 | 881 | US-09-039-555B-16 | Sequence 16, Appl1 |
| 37 | 85 | 5.2 | 4544 | US-08-469-486-52 | Sequence 52, Appl1 |
| 38 | 85 | 5.2 | 4544 | US-08-469-558-52 | Sequence 52, Appl1 |
| 39 | 84.5 | 5.2 | 2628 | US-08-570-311-14 | Sequence 14, Appl1 |
| 40 | 84 | 5.2 | 3724 | US-08-804-227C-10 | Sequence 10, Appl1 |
| 41 | 84 | 5.2 | 3724 | US-08-804-198-4 | Sequence 4, Appl1 |
| 42 | 83 | 5.1 | 514 | US-08-600-656-3 | Sequence 3, Appl1 |
| 43 | 83 | 5.1 | 531 | US-08-688-988-39 | Sequence 39, Appl1 |
| 44 | 82 | 5.0 | 455 | US-08-870-827-3 | Sequence 3, Appl1 |
| 45 | 81.5 | 5.0 | 557 | US-08-793-229-33 | Sequence 33, Appl1 |

ALIGNMENTS

RESULT 1
US-08-713-298B-2
Sequence 2, Application US/08713298B
Patent No. 5922586
GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dammann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Sch lein, Martin
APPLICANT: J rgensen, Per
APPLICANT: Bjcernevad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
TITLE OF INVENTION: Cellulytic Enzymes
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5922586 No. 5922586disk of No. 5922586th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,298B
FILING DATE: 13-SEPT-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valetta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3794,424-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 867-0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-298B-2

Query Match 12.3% Score 201; DB 2; Length 400;
Best Local Similarity 26.6% Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

```

0Y      8 VSGTLLYDANPNFPMRGI -NHGAMV-----KQATTALEGIANGANVRI 54
Db      40 ISNGELVNERGEQVQLCKMSSHGQLMGQFNPYESMKNLR-----DMGINFRA 89
QY      55 VL--SDGQWTKDDI-HTVRNLISLADENHLVAPEVHDATGYDSIASLNRAVDYTEMR 111
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      90 AMYTSGGYIDDPVKEKFKVAEALDLQIYVILIDWHILSDNPNTYKEAKDFPEEM 148
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      112 SALTGKEDTVLIINIAEMFGSMEDDAMADYK----CAIPRLRAGNHLMLVDAAQMG 167
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      149 SELIGDIDYNYIYEIAENNGS--DYTMGNQIKPYAAEYVILPIRNNDPNNIIV--GTGT 203
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      168 FPOSIIHDYGREYFNADPORMTFSLHMIETVAGNAGSAQVRYTINDEYLNODIALYIGEGHR 227
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      204 MSQDVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLRDQVDYVALDQGAALFVSENGTS 256
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      228 HTNGD---VDERTI-MSYSFGRGVGLAWS 253
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      259 AATGGGQGVFLIDEAQWVIDEMERNLSWANWS 289

```

RESULT 2
US-08-870-180B-2
; Sequence 2, Application US/08870180B

GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dandmann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Blsg rd-Frantzen, Henrik
APPLICANT: Sch lein, Martin
APPLICANT: J rgensen, Per
APPLICANT: Bloeruvad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
TITLE OF INVENTION: Cellulytic Enzymes
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59453270 No. 5945327disk of No. 5945327th America, Inc
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,180B
FILING DATE: 6-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794.434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-180B-2

Query Match 12.3%; Score 201; DB 2; Length 400;
Best Local Similarity 26.6%; Pred. No. 2e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14.

```

Db      40  ISNGELVNERBEOYQKGMSSHGQWGVNYESMKRLD-----DWGINVFR  89
QY      55  VL--SDGQMTKDI-HTVRNLISLADNHLVAVPEVHADATGDSIASLRAYDIWTEMR  111
Db      90  AMYTSSGGYIDDPVSKEKYEKVAFAALDILYIIDHILISDDNPNTYKEAADFDEM-  148
QY     112  SALIGKETVLIINIANEFESWEGEDAAVADGYK---CAIFELRNLNHTLNAVDAWGQ  167
Db     149  SELGDEYPNVYIEIENANPNS--DVTGNGOIKPFAEEVPIIRINPNNTIIV---GTGT  203
QY     168  FPOSIDHVGREVNADPORNMTFSIHXYEAGNAGSOVFNTRIDRVLNODLALYIGFGR  227
Db     204  WSDVH-HAANDQGLADP--NMYAFH--YAGTHGQVLROYVDALDQGAITVSENGTS  258
QY     228  HTNGD---VDEATI-MSYSRQGVGLANS  253
Db     259  AATGGGVLEDAQWVDFIDFDERLSTVANS  289

```

RESULT 3
US-08-814-052-4
; Sequence 4, Application US/08814052
; Patent No. 6015783

1 APPLICANT: von der Osten, Claus
2 APPLICANT: Cherry, Joel R.
3 APPLICANT: Bjornvad, Mads E.
4 APPLICANT: Vind, Jesper
5 APPLICANT: Rasmussen, Michael Dolberg
6 TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
7 TITLE OF INVENTION: OR STAINS FROM CELLULOSE FABRIC
8 NUMBER OF SEQUENCES: 55
9
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: No. 60157830 No. 6015783disk of No. 6015783th America, Inc
12 STREET: 405 Lexington Avenue, Suite 6400
13 CITY: New York
14 STATE: New York
15 COUNTRY: U.S.A.
16 ZIP: 10174-6401
17
18 COMPUTER READABLE FORM:
19
20 MEDIUM TYPE: Diskette
21
22 COMPUTER: IBM Compatible
23
24 OPERATING SYSTEM: DOS
25
26 SOFTWARE: FastSeq for Windows Version 2.0
27
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/814,052
30 FILING DATE: 06-MAR-1997
31
32 CLASSIFICATION: 510
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Lambiris, Elias J
36 REGISTRATION NUMBER: 33,728
37
38 REFERENCE/DOCKET NUMBER: 4684.204-US
39
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: 212-867-0123
42 TELEFAX: 212-878-9655
43
44 TELEX:
45
46 INFORMATION FOR SEQ. ID NO.: 4:
47
48 SEQUENCE CHARACTERISTICS:
49 LENGTH: 400 amino acids
50 TYPE: amino acid
51 STRANDEDNESS: single
52 TOPOLOGY: linear
53
54 MOLECULE TYPE: protein
55
56 US-08-814-052-4

| | | | | |
|-----------------------|-------|-----------------------------------------------------|------|------------------------------------|
| Query Match | 12.3% | Score 201 | DB 3 | Length 400 |
| Best Local Similarity | 26.6% | Pred. No. 2e-11 | | |
| Matches | 72 | Conservative | 44 | Mismatches 109; Indels 46; Gaps 14 |
| QY | 8 | VSCTLLIYANGNPFYMKRI-NHGAWY-----KQATTAIGSIANTGANTYRI | 54 | |
| | : | : | : | : |

DB 40 ISNGELVNERGEVOYOLKGMSSHGLQWYGFVNYESMKMLND-----DWGINVFR 89
QY 55 VL--SDGGQWTKDDI-HTVRNLISLAEDNHLVAVPEVHDATGDSIASLRANDVIEMR 111
DB 90 AMTSSGGYIDDPSEVKEKKEAVEAIDDIYIIMHLSNDPPIYKEADPFDEM-148
QY 112 SALIGREDYIINIANEMFGWEGMADQYK---QALPRLNAGLNHTLWDAAGWG 167
DB 149 SELGDPYVITYEIANEPGNS--DVTWGNQIKRYAEVPIIINNPNNTIIV---GTGT 203
QY 168 FPOSIDYGREVFNDPQNTMESIMTEYAGGNSQVNTNDVYNODLALVIGEGHR 227
DB 204 WSQDVA-HAADNQLADP--NVMTAFHF--YAGTHGONLRDQVYALDQGAALFVSEMGTS 258
QY 228 HTNGD----VDEATI-MSYSEORGWGLMS 253
DB 259 AATGCGVFLDEAQAVIDFMDERNLMSMANS 289

RESULT 4
US-08-812-829-4
Sequence 4, Application US/08812829
Patent No. 6017751

GENERAL INFORMATION:
APPLICANT: von der Osten, Claus
APPLICANT: Bjornvad, Mads E.
APPLICANT: Vind, Jesper
APPLICANT: Rasmussen, Michael Dolberg
TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
TITLE OF INVENTION: OR STAINS FROM CELLULOSIC FABRIC
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6017751st of No. 6017751th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,829
FILING DATE: 06-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4690.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655

TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-812-829-4

Query Match 12.3%; Score 201; DB 3; Length 400;
Best Local Similarity 26.6%; Pred. No. 26-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLYDANGNDFVNRGI-NHGHANY-----KQATTAIEGIANTGANTVRI 54
DB 40 ISNGELVNERGEVOYOLKGMSSHGLQWYGFVNYESMKMLND-----DWGINVFR 89

QY 55 VL--SDGGQWTKDDI-HTVRNLISLAEDNHLVAVPEVHDATGDSIASLRANDVIEMR 111
DB 90 AMTSSGGYIDDPSEVKEKKEAVEAIDDIYIIMHLSNDPPIYKEADPFDEM-148
QY 112 SALIGREDYIINIANEMFGWEGMADQYK---QALPRLNAGLNHTLWDAAGWG 167
DB 149 SELGDPYVITYEIANEPGNS--DVTWGNQIKRYAEVPIIINNPNNTIIV---GTGT 203
QY 168 FPOSIDYGREVFNDPQNTMESIMTEYAGGNSQVNTNDVYNODLALVIGEGHR 227
DB 204 WSQDVA-HAADNQLADP--NVMTAFHF--YAGTHGONLRDQVYALDQGAALFVSEMGTS 258
QY 228 HTNGD----VDEATI-MSYSEORGWGLMS 253
DB 259 AATGCGVFLDEAQAVIDFMDERNLMSMANS 289

RESULT 5
US-08-870-180B-13
Sequence 13, Application US/08870180B
Patent No. 5945327

GENERAL INFORMATION:
APPLICANT: Outtrup, Helle
APPLICANT: Dabmann, Claus
APPLICANT: Olsen, Arne
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Sch Jeln, Martin
APPLICANT: J rgensen, Per
TITLE OF INVENTION: DNA Constructs and Methods of Producing
TITLE OF INVENTION: Cellulytic Enzymes
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5945327th of No. 5945327th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,180B
FILING DATE: 6-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794.434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-180B-13

Query Match 12.3%; Score 201; DB 2; Length 462;
Best Local Similarity 26.6%; Pred. No. 2-5e-11;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLYDANGNDFVNRGI-NHGHANY-----KQATTAIEGIANTGANTVRI 54
DB 40 ISNGELVNERGEVOYOLKGMSSHGLQWYGFVNYESMKMLND-----DWGINVFR 89
QY 55 VL--SDGGQWTKDDI-HTVRNLISLAEDNHLVAVPEVHDATGDSIASLRANDVIEMR 111

Db 90 AMTSSGGYIDDPVSKKEVEAEADLDIYIIDHIIISDNDPNRYEADDFPEM- 148
QY 112 SALIGKEDYIINIANFWGSGWEDANADYK---QAIPLRNAGINTHTLWDAAGWG 167
Db 149 SELTGDYPPNYIYEIANEPNS--DVTWGNQIKPFAEEVPIIRNDNNIITIV---GTGT 203
QY 168 FPOSIDHGREVFNADQQRMTFSIHMEYAGGASQVFNINIDRVINODLALYIGEGHR 227
Db 204 WSDQVH-HAADNOLADP--NMYAFHF--YAGTHGQNLROQVADLQGAIFVSEWGT 258
QY 228 HINGD---VDEATI-MSYSEORGVGWLANS 253
Db 259 AATGDGGVFLEDAQVWIDFDERNLSTWANN 289

RESULT 6
US-08-727-548-2
; Sequence 2, Application US/08727548
; Patent No. 5856165
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: No. 5856165el Alkaline Cellulase and Methods of
; TITLE OF INVENTION: Producing Same
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,548
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 467 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-727-548-2

Query Match 11.7%; Score 191; DB 2; Length 467;
Best Local Similarity 24.6%; Pred. No. 2,3e-10;
Matches 69; Conservative 49; Mismatches 98; Indels 64; Gaps 14;
QY 8 VSGTTLTDANGNPFVNRGCT-NHGHAWTKDQATRAISGIANFGANTVRIYLSGGQWTKDD 66
Db 40 ISNGELVNERGEQVQLKSGMSHGLQWYG-----QFVYSEMKWLRDD 81
QY 67 -----IHT-----VRNLISLAEDNHLVAVPEVDANGYDSIASLNR 103
Db 82 WGTIVRAAMTYSSGGYIDDPVSKKEVEAEADLDIYIIDHIIISDNDPNRYEADDFPEM- 141
QY 104 VDIWIEKRSALIKEDYIINIANFWGSGWEDANADYK---QAIPLRNAGINTHTLWDAAGWG 158
Db 142 KDFDEK-SELYGDYPPNYIYEIANEPNS--DVTWGNQIKPFAEEVPIIRNDNNIITIV---GTGT 197
QY 159 MYDAAGGQFPOSIDHGREVFNADQQRMTFSIHMEYAGGASQVFNINIDRVINODLALYIGEGHR 218
Db 198 IV---GTGTWSDQVH-HAADNOLADP--NMYAFHF--YAGTHGQNLROQVADLQGAIFVSEWGT 258
QY 219 LVIGEGHRHTNGD---VDEATI-MSYSEORGVGWLANS 253
Db 250 IFVSEWGTSAATGDGGVFLEDAQVWIDFDERNLSTWANN 289

RESULT 7
US-08-924-440-2
; Sequence 2, Application US/08924440
; Patent No. 5871550
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.

APPLICANT: Frits et al.
; TITLE OF INVENTION: MUTANT THERMONOSPORA SPP. CELLULASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,440
; FILING DATE: August 27, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Christopher L. Stone
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 846-7555
; TELEFAX: (650) 845-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-924-440-2

Query Match 10.5%; Score 171.5; DB 2; Length 430;
Best Local Similarity 24.4%; Pred. No. 1.5e-08;
Matches 77; Conservative 49; Mismatches 122; Indels 67; Gaps 18;
QY 8 VSGTTLTDANGNPFVNRGCT-NHGHAWTKDQATRAISGIANFGANTVRIYLSGGQWTKDD 66
Db 137 VGGICDCDEHGNVQLKSGMSHGLQWYG-----QFVYSEMKWLRDD 81
QY 62 WT-----KDIHIVRNLISLAEDNHLVAVPEVDANGYDSIASLNRADVWIEKRSALIGK 117
Db 197 TNPGRFTDPMH---OLIMATARGLYVIYDWHILTPGDPHYNLDRAKTFFFAIAQRHASK 253
QY 118 EDTVITINIANFWGSGWEDANADYK---QAIPLRNAGINTHTLWDAAGWGQFPOSIDH 174
Db 254 TN-VLKEIANE-----PNGVSNASISYAEVYIPVIRQDPDSVIIVGIRGMSLQVS-EG 307
QY 175 YGREVFNDP--QRTMTFSIHMEYAGGASQVFNINIDRVINODLALYIGEGHRHT 229
Db 308 SGPAAEIAAPVNASIMYAFHF-----ASHRDNVYNALREASELPPVFTETGETY 361
QY 230 NND-----VDEATIMGSEORGVGWLANSWK-----GNQPEWETLDS 267
Db 362 TSDGANDPQMDR--YIDLMEKRIKIGTKWNSDDFSGAVFQPGTCASGP----- 411
QY 268 NDMAANNLTAMGNTI 282
Db 412 --WSSGSLKASGQW 424

RESULT 8
US-08-604-913B-11
; Sequence 11, Application US/08604913B
; Patent No. 5712142
; GENERAL INFORMATION:
; APPLICANT: Adney, William S.
; APPLICANT: Thomas, Steven R.

APPLICANT: Himmel, Michael E.
APPLICANT: Baker, John O.
TITLE OF INVENTION: METHOD FOR INCREASING
TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: U.S.A.
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASC II (DOS) text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/604,913B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,213
FILING DATE: 15-07-1994
ATTORNEY/AGENT INFORMATION:
NAME: Edna M. O'Connor
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: 95-56
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/384-7573
TELEFAX: 303/384-7499
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
FEATURE:
NAME/KEY: E1-CAT
US-08-604-913B-11

Query Match 8.7%; Score 142; DB 1; Length 358;
Best Local Similarity 23.1%; Pred. No. 7.8e-06;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 2 ANSGF-YVSGTLYDANGNPFVARGIN-----HGHWKDKQATTAIEBIANTGA 49
DB 1 AGGGYWTSGREIIDANNVPRAGINWFGFETCNVYHG-LWSRDY-RSMLDQIKSLGY 58
QY 50 NTVRIYVSDSGOWTKDID---HTVRNLISLAEDNHLVAVPEVHATGYDSIASINRAVDY 106
DB 59 NTRILRPLYS-----DILRPGITPNSINFYQNN-----QDLQGLTSLQVYDKIVAY 103
QY 107 -----WIENRSALI-----GKEDVTIINIANEW 129
DB 104 AGOIGLRILIDRRHPCSGOSALWYTSVSEATWISDLQALQRYKGNPVTYVGFDLNE- 162
QY 130 FGSWEGAMADG-----YKQAIPLRNA--GLNHTLAV-----DAAGW----- 165
DB 163 --PHDPACWCCGDPISIDWRLAERAGNAVLNVNPNLLIFEGVQSYNGDSYWGNGNLOGA 220
QY 166 GQFP-----QSITHDGREVFENDPQNTWFSIHMYEYAGNNSQV-RTNIDRYLN 214
DB 221 GQFPVVLNVNRLVYSADHATISY---PQ--TWFSPTFP---NNNPGIMKNKMGILFN 272
QY 215 QDLALV-IGERGH--RTNGDVDEATIMSY---SEQRGVG---WLANSWGNG----- 258
DB 273 QNLAPVWIGEGEGLTQSTDTQWTKLTVQYLRPTAQYGAUSFWTFWMSWNPDSGDTGGIL 332
QY 259 -PEWEYLDLSND 269

DB 333 KDDWQIVDVKD 344

RESULT 9
US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 8.7%; Score 142; DB 1; Length 521;
Best Local Similarity 23.1%; Pred. No. 1.3e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 2 ANSGF-YVSGTLYDANGNPFVARGIN-----HGHWKDKQATTAIEBIANTGA 49
DB 1 AGGGYWTSGREIIDANNVPRAGINWFGFETCNVYHG-LWSRDY-RSMLDQIKSLGY 58
QY 50 NTVRIYVSDSGOWTKDID---HTVRNLISLAEDNHLVAVPEVHATGYDSIASINRAVDY 106
DB 59 NTRILRPLYS-----DILRPGITPNSINFYQNN-----QDLQGLTSLQVYDKIVAY 103
QY 107 -----WIENRSALI-----GKEDVTIINIANEW 129
DB 104 AGOIGLRILIDRRHPCSGOSALWYTSVSEATWISDLQALQRYKGNPVTYVGFDLNE- 162
QY 130 FGSWEGAMADG-----YKQAIPLRNA--GLNHTLAV-----DAAGW----- 165
DB 163 --PHDPACWCCGDPISIDWRLAERAGNAVLNVNPNLLIFEGVQSYNGDSYWGNGNLOGA 220
QY 166 GQFP-----QSITHDGREVFENDPQNTWFSIHMYEYAGNNSQV-RTNIDRYLN 214
DB 221 GQFPVVLNVNRLVYSADHATISY---PQ--TWFSPTFP---NNNPGIMKNKMGILFN 272

Db 44 VQGNOLVQSGQANQVLVSMSSHGLWYCN-----FYKSSGLQWMDN 85
Qy 67 --HTVNLILAEEDNLVAVPEVDATGYSIASLNA-----VDYWT----- 108
Db 86 WGINFRAAMTSEBGX-ITDPSYKKNKVEANQASIDLALYIIDLWMLSDGNPTYKAQ 144
Qy 109 -----EMASALIGKEDYIINIANEMFSGWGDMAADGKALPRLRNAGLNHTLVDA 163
Db 145 SKAFQENATLYGNTPNVITYELATSPTECVLGRQSS--EEVITAIRSIDPQGVIVGSP 202
Qy 164 GMGCFPOSIDHYGREVFENADPQRTMFSIMHYEYAGNAGVYRINDRYLNODLALVIGE 223
Db 203 TWSQ-----DHLAADNPVSHSNMYALHF--YSGTHGQPLRDXITTAAMKGAALFYTE 254
Qy 224 FGHRTNGD---VDEATIMSYSEQRGVGLAWS 253
Db 255 WGTSDASGNGPYLPQSKEMIDFLNARKISWYMS 289

RESULT 12
US-08-745-977-4
; Sequence 4, Application US/08745977
; Patent No. 5770406
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglycanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5770406 No. 5770406disk of No. 5770406th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745,977
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4175,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-745-977-4

Query Match 6.1%; Score 100; DB 1; Length 429;
Best Local Similarity 21.9%; Pred. No. 0.11; Indels 88; Gaps 18;
Matches 68; Conservative 35; Mismatches 119; Indels 88; Gaps 18;
Qy 27 NHGAWTKDQATTAIEGIANTGANTVRIVLSGGQWTKDIDHTVYN----- 72
Db 106 NHYDWIN---PATVQSHVDGLMTIRIP---GYWSNALVDNASEFFADGNQLPFLD 159

Qy 73 -LISAEENELVAVPEVDATG---YDSIASLN-----RAVDY-----WIEMSA 113
Db 160 AVYORRADLGIYIYIIDLGAAGGQODAFGTGONPNAGFYNTYYGRAEKLMSWTNIIH 219
Qy 114 LIGKEDY--IINIANEMFSGWGDMA--ADG-----YKQAIPLR-----NAGLN 155
Db 220 TNPASTYGVKLEVLNEPYSRBDGGGRYRAPGODDSMVOTYIFGALKAVRDAEALNVP SN 279
Qy 156 HTLWDAAGKQFQPSIDHYGREVFENADPQRTMFSIMHYEYAG-----GNASQVR--- 206
Db 280 KKLHY-----QFMSKMDSGDPSRNAVKNDPVWGFDHNYIGFALQNTGDQSLMBSA 333
Qy 207 -TNDRYLNODLALYIGFGRHTNG--DYDQATIMS-----YSEQRGVGLAWSK 255
Db 334 CDSRVVSGQDFA-ITGEMS--MTSGADWHDGNFTKFTAQOQLYESPGMDGWYWTWK 390
Qy 256 G--NGEMEX 263
Db 391 TELNDPWTY 400

RESULT 13
US-09-040-699A-4
; Sequence 4, Application US/09040699A
; Patent No. 6022723
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Andersen, Lene N.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Dalboge, Henrik
; APPLICANT: Olsen, Hans S.
; TITLE OF INVENTION: An Enzyme With Beta-(1-6)-Endoglycanase
; TITLE OF INVENTION: Activity
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 6022723 No. 6022723disk of No. 6022723th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,699A
; FILING DATE: 18-March-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4175,214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-040-699A-4

Query Match 6.1%; Score 100; DB 3; Length 429;
Best Local Similarity 21.9%; Pred. No. 0.11; Indels 88; Gaps 16;
Matches 68; Conservative 35; Mismatches 119; Indels 88; Gaps 16;
Qy 27 NHGAWTKDQATTAIEGIANTGANTVRIVLSGGQWTKDIDHTVYN----- 72

Db 106 NHYEDWIN---PATVQSVHVDGLNTIRPI---GYMSYNAIVDTASPEADGNLTOLPYLD 159
QY 73 -LISLAEDNHLVAVPEVHDATG---YDSIASLN-----RAYDY-----WIEBBSA 113
Db 160 AVVOKAADLGYIVITIDLGAAGGGOODAFITGONPNPAGFYNTYDYGRAEKWLSMTNINIH 219
QY 114 LIGKEDVY-IINIANEMFGSWGDAM--ADG-----YKQAIPLRL-----NAGLN 155
Db 220 TNPATSTGMEVLENEPVSRRDGGGRYAPGQDPSMGTYYTPGALKAVRDALALNPSPN 279
QY 156 HTLWADAAGWGFQPSIHIDYGEVFNADPQNTMFSIHMTIAG-----GNASQVR--- 206
Db 280 KKLHV-----QFMSSKMDSGDPRSNAAVKNDDPMVGFDDHNTIGFALQNTGDOYSLHSA 333
QY 207 -TNDIVYANODLALYIGFEGHRTNG-DVDEATIMS-----YSHQGVGLWAMSK 255
Db 334 CTDSRVVSGQDFA-TTGEMS--MTISGALWHDGNEFTTKFTTAQOOLYESPGMDGWTYTWK 390
QY 256 G--NGPEWEY 263
Db 391 TELNDPRWTY 400

RESULT 14
US-09-066-075-2
Sequence 2, Application US/09066075
Patent No. 5925749

GENERAL INFORMATION:

APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,075
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-066-075-2

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

QY 19 NPFVAMGINSNGHA-----W---YKDAQATRAIEGANTGANTVRIYVLSDDGOWTK--- 64
Db 9 NKLIGRINIGNALLEAPNEDGQGVVTKDE---FFDIKRGSSHVPII---RMSHAY 61
QY 65 -----DDI--HTVRNLISLAEDNR-----LVAVPEVDATGYDSIASLNRA 103
Db 62 AEPPIKIMDFEKKRVDEVINGALKRGLAVAINHHEELMNDPEERK-----ERF 111
QY 104 VDYWIEKRSALIGKEDVTIINIANEMFGSWGDAMADYKQAIPLRLNGLNHTLWADA 163
Db 112 LALMKQADRYKDYPELTFEELINEPHGNLTPEKMWELLEALKVRSIDKRTIIGTA 171
QY 164 GWGQFPQ---SHDYGREVFNADPQNTMFSIHMT---EYAGNMQVNTNIDRYLNOD 216
Db 172 EWGGSIALEKISVPKW-----EKNSIVTIIHYNPFEFTHOGAEWYEGS--EKLGRK 221
QY 217 LA-----LVYIGFEGHRTNGVDV-----ATMSYSEQRC 246
Db 222 WGSDDQKHLIEENFEIEMSKKKNRIYIGERG-AVRKADLSRLKWTSEVYREMEKR- 279
QY 247 VGWLAMSKNGNGPEWEY 263
Db 280 ---RMSW---AYWEF 288

RESULT 15
US-08-518-615A-2
Sequence 2, Application US/08518615A
Patent No. 5962258

GENERAL INFORMATION:

APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615A
FILING DATE: August 23, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-518-615A-2

Query Match 6.1%; Score 99.5; DB 2; Length 317;
Best Local Similarity 18.9%; Pred. No. 0.077;
Matches 60; Conservative 54; Mismatches 94; Indels 109; Gaps 17;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:10:59 ; Search time 245.85 Seconds
(without alignments)
77.438 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330

Perfect score: 1630
Sequence: 1 NANSGFYVSGTLVDANGNP.....TIVNGPYGLRHSRLSTVPT 300

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.65:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|----------------------|
| 1 | 1169 | 71.7 | 516 | 2 JE0134 | mannan endo-1,4-beta |
| 2 | 710 | 43.6 | 384 | 2 T37232 | secreted beta-mann |
| 3 | 692 | 42.5 | 363 | 2 S30386 | mannan endo-1,4-beta |
| 4 | 601 | 36.9 | 1331 | 2 A48954 | mannan endo-1,4-beta |
| 5 | 200 | 12.4 | 409 | 2 B25156 | cellulase (EC 3.2. |
| 6 | 200 | 12.3 | 488 | 2 A25156 | cellulase (EC 3.2. |
| 7 | 197 | 12.1 | 557 | 2 JC5487 | cellulase (EC 3.2. |
| 8 | 184.5 | 11.3 | 505 | 2 S39962 | endoglucanase - Er |
| 9 | 180 | 11.0 | 448 | 2 A27631 | cellulase (EC 3.2. |
| 10 | 176.5 | 10.8 | 466 | 2 C42360 | cellulase (EC 3.2. |
| 11 | 174.5 | 10.7 | 499 | 2 A27198 | cellulase (EC 3.2. |
| 12 | 174.5 | 10.7 | 504 | 2 S54744 | cellulase (EC 3.2. |
| 13 | 173.5 | 10.6 | 508 | 2 A26874 | cellulase (EC 3.2. |
| 14 | 162.5 | 10.0 | 486 | 2 I40548 | bifunctional cellu |
| 15 | 162.5 | 10.0 | 508 | 2 G69593 | cellulase (EC 3.2. |
| 16 | 157 | 9.6 | 783 | 2 JC5467 | cellulase (EC 3.2. |
| 17 | 152 | 9.3 | 800 | 2 A29003 | cellulase (EC 3.2. |
| 18 | 152 | 9.3 | 822 | 2 JT0611 | cellulase (EC 3.2. |
| 19 | 151.5 | 9.3 | 825 | 2 JN0114 | cellulase (EC 3.2. |
| 20 | 151.5 | 9.3 | 825 | 2 JS0114 | cellulase (EC 3.2. |
| 21 | 149.5 | 9.2 | 357 | 2 PC4404 | cellulase (EC 3.2. |
| 22 | 149.5 | 9.2 | 941 | 2 S29043 | cellulase (EC 3.2. |
| 23 | 147.5 | 9.0 | 429 | 2 S29044 | endoglucanase A pr |
| 24 | 136.5 | 8.4 | 570 | 2 S56132 | cellulase (EC 3.2. |
| 25 | 131 | 8.0 | 814 | 1 CZGLDM | cellulase (EC 3.2. |
| 26 | 130 | 8.0 | 426 | 2 A42649 | cellulase (EC 3.2. |
| 27 | 127 | 7.8 | 915 | 2 PC4278 | guar gum-degrading |
| 28 | 122.5 | 7.5 | 915 | 2 A43802 | cellulase (EC 3.2. |
| 29 | 122.5 | 7.5 | 1039 | 2 S02711 | cellulase (EC 3.2. |

| | | | | | |
|----|-------|-----|-----|----------|--------------------|
| 30 | 122 | 7.5 | 438 | 2 A47702 | glucan 1,3-beta-gl |
| 31 | 117.5 | 7.2 | 611 | 2 S03767 | cellulase (EC 3.2. |
| 32 | 116 | 7.1 | 428 | 2 JC7177 | endoglucanase V (E |
| 33 | 115.5 | 7.1 | 747 | 2 B47093 | cellulase (EC 3.2. |
| 34 | 114 | 7.0 | 419 | 2 S72325 | glucan 1,3-beta-gl |
| 35 | 113.5 | 6.8 | 517 | 2 I40798 | cellulase (EC 3.2. |
| 36 | 111.5 | 6.8 | 441 | 2 A44815 | cellulase (EC 3.2. |
| 37 | 111 | 6.8 | 475 | 1 CZGLCA | cellulase (EC 3.2. |
| 38 | 110 | 6.7 | 26 | 2 PC4285 | guar gum-degrading |
| 39 | 109.5 | 6.7 | 734 | 2 T14877 | hypothetical prote |
| 40 | 107 | 6.6 | 430 | 2 S55325 | endo-beta-1,6-gluc |
| 41 | 105 | 6.4 | 364 | 2 S12017 | endoglucanase A - |
| 42 | 105 | 6.4 | 456 | 2 T40276 | probable gas1 fam1 |
| 43 | 104 | 6.4 | 556 | 2 A40589 | cellulase (EC 3.2. |
| 44 | 101.5 | 6.2 | 317 | 2 B72216 | endoglucanase - Th |
| 45 | 100.5 | 6.2 | 748 | 2 S19652 | cellodextrinase C |

ALIGNMENTS

RESULT 1
JE0134
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Bacillus circulans
N: Alternate names: endo-1,4-beta-mannanase
C: Species: Bacillus circulans
C: Date: 03-Jun-1998 #sequence-revision 10-Jul-1998 #text-change 17-Mar-1999
C: Accession: JE0134
R: Yoshida, S.; Sako, Y.; Uchida, A.
Biosci. Biotechnol. Biochem. 62, 514-520, 1998
A: Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene c...
A: Reference number: JE0134; MUID:98233274
A: Accession: JE0134
A: Molecule type: mRNA
A: Residues: 1-516 <YOS>
A: Cross-references: DDBJ:AB007123
A: Note: the authors translated the codon CAA for residue 259 as Asn. CAA for residue
A: Note: the translation of the nucleotide sequence 1294-1365 is not given in this pap
C: Keywords: glycoprotein; glycosidase; hydrolase

| Query Match | Score | DB 2: | Length | 516: |
|------------------------------------------------------------------------------|------------------|------------------------------------------------------------|---------------------------------------|---------|
| Best Local Similarity | 69.2% | Pred. No. 1e-83: | | |
| Matches 207: | Conservative 46: | Mismatches 44: | Indels 0: | Gaps 0: |
| QY | 1 | NANSGFYVSGTLVDANGNP | RVKRGINHGAMWYDQATTAEGTANTGANTVRIYLSDG | 60 |
| DB | 33 | HAASGFYVSGTKLDANGOPFVARGVNAHWKQDLSAIPATAKIGANTIRIVLANGH | 92 | |
| QY | 61 | QWTDDHTVNLISLADNHLVAVPEYHATGYSIASLNRAVDYWTMRSAIIGKEDT | 120 | |
| DB | 93 | KWTLDDVTVNNITLCEONKTLIAVLEVHDATGSDLSIDLNAVYWGIRKSALIGKEDR | 152 | |
| QY | 121 | VVINANWESGWDADGDKOAIPELRNAGNHTLVDAAGMGPDSIHVGREVF | 180 | |
| DB | 153 | VVINANWESGWDADGDKOAIPELRNAGNHTLVDAAGMGPDSIHVGREVF | 212 | |
| QY | 181 | NADPQNTFISIHMYEYAGNACVTRINIDRVLMODLALVIGFGRHTNGDVDEATIMS | 240 | |
| DB | 213 | NADPLKNTVFISIHMYEYAGNACVTRINIDRVLMODLALVIGFGRHTNGDVDEATIMS | 272 | |
| QY | 241 | YSEGRGVGLAMSWKGNSEYDLSNDAGNNLTAVGNTIVNGPYGLRHSRLSTVPT | 299 | |
| DB | 273 | YSNEGVGLAMSWKGNSEYDLSNDAGNNLTAVGNTIVNGPYGLRHSRLSTVPT | 331 | |
| RESULT 2 | | | | |
| T37232 | | | | |
| secreted beta-mannosidase - Streptomyces coelicolor | | | | |
| C: Species: Streptomyces coelicolor | | | | |
| C: Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 03-Dec-1999 | | | | |
| C: Accession: T37232 | | | | |
| R: Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rørdam, M.A. | | | | |

Db 234 DLAAD-----NPINDSNMYSVHF--YSGTNPISVDFNRDANMSNRYALNHGAAYFA 285
 QY 222 GEFGRHRTNGDVDEATIMSYSEORGVGLAMSKNGKNGPEWEYIDLND----- 269
 Db 286 TEMGTSLATG-----TTGPTL-AKADAWLDF-LNGNNISWNCFSISNKDEKAALNLSLS 338
 QY 270 -----WAGNNLTAMGNTI---VNGPY 287
 Db 339 LDPGSDKLMADNELTTSQGVYRARIKQAY 367

RESULT 8
 S39962
 C:Species: Erwina carotovora
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
 C:Accession: S39962
 R:Cooper, V.J.C.; Salmond, G.P.C.
 Mol. Gen. Genet. 241, 341-350, 1993
 A:Title: Molecular analysis of the major cellulase (celV) of Erwina carotovora: evidenc
 A:Reference number: S39962; MID:94067016
 A:Accession: S39962
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-505 <COO>
 A:Cross-references: EMBL:X76000; NID:g434941; PIDN:CAA53592.1; PID:g434942

Query Match 11.3%; Score 184.5; DB 2; Length 505;
 Best Local Similarity 23.1%; Pred. No. 6.4e-07;
 Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

QY 13 LVDANNPVFMGI-NHGAHYDQATATAGTANGANTVRLVSDGGWTD----- 66
 Db 48 LVDEGKRVQLRGISSHGLWFGD-----YNNDSMKMTDMDGIV 89
 QY 67 -----IHTVNLISLADNHLVAVPEVHDATGYDSIASLRAYDYNT 108
 Db 90 FRVAMTADGYSNPSLNKVEVAAMASLGVYIIIDPHILSDNDPNYKQAKTFRA 149
 QY 109 EKRSLIGKEDVYIINANEWFG-SWEGD-AMADGYKQALPRLNAGLNTIYAVDAAG 164
 Db 150 EM-AGLYGSSPNYIETIANEPNGVWNGQIRPYA---LEVDTJISKDPDNLIY---G 202
 QY 165 WQGFPSIDHYGRVFNADPQRTMSIHETFAAGNASVYRNIDRYLQDLALVIGEF 224
 Db 203 TGTWSDIDH--AADNQLPDPMTMALHP--YAGTHGQFLRDIRIDYASRGAIIVSEW 257
 QY 225 GHRHTNGD-----VDEATIMSYSEORGVGLAMSKNGKNGPEWEYL---DLSNDWAGNNL 275
 Db 258 GTSDSAGNGPFLPESQWTWIDFLNNGVSWVWNSLIDKSEASALAPGASKSGGWTBQNL 317
 QY 276 TAMGNTI 282
 Db 318 STSGKEV 324

RESULT 9
 A27631
 C:Species: Clostridium acetobutylicum
 N:Alternate names: endo-1,4-beta-glucanase
 C:Date: 31-Dec-1988 #sequence_revision 30-Jun-1991 #text_change 15-Oct-1999
 C:Accession: A27631
 R:Zappe, H.; Jones, W.A.; Jones, D.T.; Woods, D.R.
 Appl. Environ. Microbiol. 54, 1269-1292, 1988
 A:Title: Structure of an endo-beta-1,4-glucanase gene from Clostridium acetobutylicum P2
 A:Reference number: A27631; MID:88268074
 A:Accession: A27631
 A:Molecule type: DNA
 A:Residues: 1-448 <ZAP>
 A:Cross-references: EMBL:M31311; NID:g144789; PIDN:AAA23230.1; PID:g144790
 A:Note: the authors translated the codon GAG for residue 116 as Gly, GAA for residue 263

C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 11.0%; Score 180; DB 2; Length 448;
 Best Local Similarity 22.1%; Pred. No. 1.2e-06;
 Matches 71; Conservative 59; Mismatches 135; Indels 56; Gaps 14;

QY 8 VSGTLLYDANGNPFVWFGI-NHGAHY-----KQDQATTAIEGIANTGANTVRI 54
 Db 49 VVGSQGLCDNSGNKPIQKSGSHGLQVYFVYVDSMKELRDK-----WGVAVIRA 98
 QY 55 VL--SDGGQMT--KDDIHTVNLISLADNHLVAVPEVHDATGYDSIASLRAYDVIEM 110
 Db 99 AMYTNEGGYISNPSQKREKIRIVODAILNVIYIIDPHILSDNDPNYKQAKSEFQEM 158
 QY 111 RSALIGKEDVYIINANEWFGSMWEGDADGYK---QAIPLRNAGLNTIYAVDAAGW 166
 Db 159 AEE-YKYSNVYIETIANEPNG--GTNWDMDIKPYANTYIPALRIDPNNTIIVGTSTWS 214
 QY 167 QFPQSIDHYGRVFNADPQRTMSIHETFAAGNASVYRNIDRYLQDLALVIGEF 224
 Db 215 QVDYDIAD-----NPLRSGNIMYTCHE--YAGTHQSLSKDKINYASKGIAIFVTEM 264
 QY 225 GHRHTNGD-----VDEA-TIMSYSEORGVGLAMSKNGKNGPEWEYL---DLSNDWAGNNL 275
 Db 265 GTSDSAGNGPFLPESQWTDPMASKNISTWMLALCDKSEASALKSGSSTGGWTDSDL 324
 QY 276 TAMGNTIYVPGYLRSTSL 296
 Db 325 TTSGLFVKKSGISNTTSOTS 345

RESULT 10
 C42360
 C:Species: Thermomonospora fusca
 N:Alternate names: endo-1,4-beta-glucanase
 C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 20-Mar-1998
 C:Accession: C42360
 R: Rao, G.; Ghanges, G.S.; Jung, E.D.; Wilson, D.B.
 J. Bacteriol. 173, 3397-3407, 1991
 A:Title: DNA sequences of three beta-1,4-endoglucanase genes from Thermomonospora fus
 A:Reference number: A42360; MID:91258320
 A:Accession: C42360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-466 <LAC>
 A:Cross-references: GB:L01577; NID:g154693; PID:g154694
 C:Function:
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A:Pathway: cellulose degradation
 C:Superfamily: bacterial cellulase-binding domain homology
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:38-137/Domain: Bacterial cellulose-binding domain homology <BCB>

Query Match 10.8%; Score 176.5; DB 2; Length 466;
 Best Local Similarity 25.0%; Pred. No. 2.4e-06;
 Matches 78; Conservative 48; Mismatches 125; Indels 61; Gaps 17;

QY 8 VSGTLLYDANGNPFVWFGI-NHGAHYDQAT-TAIEGIA-NTGANTVRI--VLSGGQW 62
 Db 173 VCGTQCDCEHGNPVQLRGKSTGIQWFDHCLDSSLDALAYDMKADITLSYIGEDGYE 232
 QY 63 TDDDIHTVA--NLISLADNHLVAVPEVHDATGYDSIASLRAYDVIEMKSLIGKEPT 120
 Db 233 TNPREFTRIDILDMATARGLYVIVDWHLPLPGDPHYVLDRAKTFEAFIAQRHASKTN- 291
 QY 121 VTIINANEWFGSMWEGDAMD--GY-KQAIPLRNAGLNTIYAVDAAGWQFQSIHDYGR 177

```

Db 292 VLYEIANE-----PNCVSNASIKSYAEVYIPVIRORDPDSVIIVGRGSSSLGVS-EGSGP 346
QY 178 EVFNADP--QKNTFESIMYETAGNAGNOVTNIDRVINODLAL--VIGFGRHRTNGD 232
Db 347 AEIAPNPNASINIMAFHFY-----AASHRDNYLNALREASELPPVYTERGETETGYGD 400
QY 233 -----VDENTIMYSQRQGVGWLANSWK-----GNGPEWEYLDLSNDW 270
Db 401 GANDFQMDR--YIDLMAERIKIGWKRMYSDDFRSGAVFQPTCASGP-----W 448
QY 271 AGNNLTANGNTI 282
Db 449 SGSSLSKASGQWY 460

RESULT 11
A27198
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain IFO3034)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 15-Oct-1999
C:Accession: A27198
R:Nakamura, A.; Dozum, T.; Beppu, T.
Eur. J. Biochem. 164, 317-320, 1987
A:Title: Nucleotide sequence of a cellulase gene of Bacillus subtilis.
A:Reference number: A27198; MUID:87190397
A:Accession: A27198
A:Molecule type: DNA
A:Residues: 1-499 <NNA>
A:Cross-references: GB:M28332; NID:g142670; PIDN:AAA22307.1; PID:g142671
A:Experimental source: strain IFO3034
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-36/Domain: signal sequence #status predicted <SIG>

Query Match 10.7%; Score 174.5; DB 2; Length 499;
Best Local Similarity 21.1%; Pred. No. 3.8e-06;
Matches 70; Conservative 66; Mismatches 123; Indels 73; Gaps 15;

QY 3 NSGFYVSTIYDANGNFVNRGT-NHGHNATKQDATTAGIANTGANTYRIYVSDGGQ 61
Db 39 NGQISLGTQLVNRDGRKAVOLKGLISSHGLQWYGD-----FYVKRDSIK 80
QY 62 WTKDD-----IHT-----VRNLISLAEDNHLVAVPEVDATGYDSIA 98
Db 81 WLRDMDGTYVRAMATYTDGTYIDNPVKYKKEVNAKEIGIYVITDWHILNDGNFQ 140
QY 99 SLNRADVIMRSALIGKEDYIINIANEFG--SWEGD--AMADGYKQAIPLRLNAGL 154
Db 141 HKEKAKFFKEM--SLYGNTPNVIYEIANEPNGDVNFKRDIKPYAE---EYISVIRKNDP 196
QY 155 NHTLNVDAAGGQFPOSIDHGREVFNDPQ--RNTFESIMYETAGNAGNOVTNIDR 211
Db 197 DNIITV--GGTWSQDND-----AADDDLKCANVYALHF--YAGTHQSLSLDRKNY 245
QY 212 VLNODLALVIEFGFGRHRTNGD--VDEA-TIMSYSEORGVGLANSKNGPEWEYL-- 264
Db 246 ALISGAPLFTVEWGTSDASGNGGYFLDOSREMLNTLDSKNTISWVWNLSDKOESSALKP 305
QY 265 --DISNDWAGNNLTANGNTIYNGPYGLRETSR 294
Db 306 GASKTGWPLTDLTAAGTFVRENILGNRDSRK 337

RESULT 12
S54744
cellulase (EC 3.2.1.4) CelY precursor - Erwinia carotovora (SCC 3193)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Erwinia carotovora
A:Variety: SCC 3193

```

```

C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S54744; S44996
R:Mae, A.; Heikkinen, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A:Title: Structure and regulation of the Erwinia carotovora subspecies carotovora SCC
A:Reference number: S54744; MUID:95231512
A:Accession: S54744
A:Molecule type: DNA
A:Residues: 1-504 <MAE>
A:Cross-references: EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g493493
C:Genetics:
A:Gene: celY1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-504/Product: cellulase #status predicted <MAT>

Query Match 10.7%; Score 174.5; DB 2; Length 504;
Best Local Similarity 23.7%; Pred. No. 3.9e-06;
Matches 73; Conservative 43; Mismatches 123; Indels 69; Gaps 14;

QY 13 LYDANGNFFVNRGT-NHGHNATKQDATTAGIANTGANTYRIYVSDGGQTKDD--IHT 69
Db 48 LVDEGKRVOLRGISNSGLQWYGD-----YVKKDSMKWLRDPMGJNV 89
QY 70 VRNLISLAEDNHLVAVPEV-----HDATGYDSISLNRADVW 107
Db 90 FRVAVTYAE-NGYINPBLANKVEAVAAOGLGYIITDHTLSDNDPNTYKQAKTFE 148
QY 108 IEMRSALIGKEDYIINIANEFGS--WEGD--AMADGYKQAIPLRLNAGNHTLNVDA 163
Db 149 AEM-AGLIGSNPNVYIEIANEPNCSVTWNGDIRPYA--LEVTDTRIRKDDNLIIVSG 204
QY 164 GWGQFPOSIDHGREVFNDPQNTMFSIMYETAGNAGNOVTNIDRVINODLALVIGE 223
Db 205 TWS--QDIHD--AADQLPDPNTLYALHF--YAGTHGQFLRDIRDYAOSRGAIFVSE 256
QY 224 FGRHRTNGD-----VDENTIMYSQRQGVGWLANSKNGPEWEYL--DLSNDWAGNN 274
Db 257 WGTSDASNGGPTLPESQTIWIDFLNNGISWVWNLSDKSETSAALVAGASKSGGWTEQN 316
QY 275 LTANGNTI 282
Db 317 LSTSGKFV 324

RESULT 13
A26874
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.
A:Reference number: A26874; MUID:87194581
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB1>
A:Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A:Note: The authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

```


A:Pathway: cellulose degradation
C:Keywords: alternative initiators; extracellular protein; glycosidase; hydrolase; polys
F:1-38/Domain: (or 2-38) signal sequence #status predicted <StG>

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 10.6%; | Score 173.5; | DB 2; | Length 508; |
| Best Local Similarity | 20.8%; | Pred. No. 4.7e-06; | | |
| Matches 69; | Conservative 68; | Mismatches 122; | Indels 73; | Gaps 15; |

```

OY      3 NSGFVSGTLLTDANGNEFNFMGI--NHGHAWTKDQATTAIEGANTGANVTAVIYLSDDGO   61
        :|::||:::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      48 NGOLSIKGTQLVNRGRKAVQVLKGISSHGLQWYGD-----FVNKKDSLK    89

OY      62 WTKDD-----IHT-----VRNLISAEDNHLVAPEVEDATGYDSIA   98
        ||::||::||::||::||::||::||::||::||::||::||::||::||
Db      90 WLRLDWGTFIVFAAAAYTTLDGGIIDNPVSKRYKEVEAKELGIYVILDMHLLNDGNPNQ   149

OY      99 SLNRAVDVIEMRSALIGKEDIYINIANENFG--SWEGD--AMADGYKOALPRLRNAGL   154
        ::||::||::||::||::||::||::||::||::||::||::||::||
Db     150 NKERAKEFFKEN--SLLYGTVPTVITIELANEPNGDVNMKKRIDIPAE--EVISVIRKNDP   205

OY     155 NHTLWAVDAWGQFPQSIIHDYGREVFNADPQ---KNTMFSIMTEYAGSNASQVATNTIDR   211
        ::||::||::||::||::||::||::||::||::||::||::||::||
Db     206 DNIITV---GCTGWSDQVDN-----AADDQLKDANVMYALTF--YAGTHGSLSDKRANY   254

OY     212 VLNDLIALVIGEGRHRTNGD---VDEA-TIMYSSEOGVGWLMSWKNGSPWEWEYL--   264
        ::||::||::||::||::||::||::||::||::||::||::||::||
Db     255 ALSRKAPFIPTVEMTGSDASNGCVFLDQGREMLNTIDSKNISVMWNMLNSDKQESSALKP   314

OY     265 ---DLSDMWAGNNLTAMGNTIYNGPYGLRETSR   294
        |||||::||::||::||::||::||::||::||::||::||::||
Db     315 GASKTGWPLDIPIASGTFRENINGTKDSTK   346

```

RESULT 14
I40548
bifunctional cellulase precursor - Bacillus sp.
C.Species: Bacillus sp.
C.Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C.Accession: I40548
R.Han, S.J.; Yoo, Y.J.; Kang, H.S.
J. Biol. Chem. 270, 26012-26019, 1995
A.Title: Characterization of a bifunctional cellulase and its structural gene: the cel
A.Reference number: I40548; MUID:96029707
A.Accession: I40548
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-486 <RES>
A.Cross-references: EMBL:U27084; NTD:g857575; PIDN:AAC3478.1; PID:g857576
C.Genetics:
A.Gene: cel

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 10.0%; | Score 162.5; | DB 2; | Length 486; |
| Best Local Similarity | 21.0%; | Pred. No. 3.2e-05; | | |
| Matches 66; | Conservative 75; | Mismatches 136; | Indels 37; | Gaps 15 |

```
QY      3 NGGFVSSGTLTDKANDNPFVMGI--NEGHAWKDQT-TAIGSIANTGANTV---RIVLS    57
Db     39 NGLSLSTGRQLNRDCKAKAVOLGTISHSHLQWGTGEYNNKDSLMLRLDRMGITFRFAMMYTA    98
QY     58 DCGOWMTKDDI-HTVRLTILSLAEDNHLVAPEVDATGYDSIASLNRAVDWTYEMRSALIG   116
Db     99 DGGYIDNPESVKKKVKAVALAKELGTYIILDMHLLINDGPNQNKKEAKKEFFKEM-SSLIG   157
QY    117 KEDVTIIINIANEFG--SWEGD--AWADGYKQAIPRLRNAGLNTHTLMVDAAGWGQFPQS I   172
Db    158 NTPNVYEIEIANEBNGDVNNKRDIKPYAE--EVIYSIRKNDPNIIITY--GRTWSODV    211
QY    173 HXYGRVFVNADPQ---RMMEFSIMHYEVYAGNASQVRINTDRLNDLNLVYGEPEHRHT    229
Db    212 ND-----AADDLKLDANWYIALHF--YAGTHGOFLFDKANXALTSBGADPIYTEWETSDA    263
```

[illegible]

RESULT 15
669593
cellulase (EC 3.2.1.4) bglc precursor - *Bacillus subtilis*
N:Alternate names: endo-1,4-beta-glucanase
""""""""""bgl-47

C:Date: 05-Dec-1997
C:Accession: 669593
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertoni, A.M.; Alloni, G.; Azevedo, V.; Bar
A.; Brown, S.; Broilletter, S.; Brusch, C.V.; Caldwell, B.; Capuano, J.; Carter, N.M.;
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Furua, S.; Galizzi, A.; Gal
lech, U.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koester, P.; Konigslehn, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lepidos, A.; Lindro
Y, M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Rieger, M.; Rivolta, C.; Roch, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scall
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
T.; Winters, P.; Wippl, A.; Tanaka, T.; Terstips, P.; Tomomi, A.; Tosato, V.; Uchida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Yano, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
A:Reference number: A69580; MUID: 98044033

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <RIND>
A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CABJ3696.1; PID:e13844
A:Experimental source: strain 168
R:MacKay, R.M.; Lo, A.; Miliute, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, ...
Nucleic Acids Res. 14, 9159-9170, 1986
A>Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
A:Reference number: A26114; MUID:8706783
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 10-508 <MAC>
A:Experimental source: strain PAP15
A>Note: part of this sequence, including the amino end of the mature form, was confirmed
R:Rindahl, V.; As, K.; Tromsø, A.
Antonie Van Leeuwenhoek 66, 327-332, 1994
A>Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtilis
A:Reference number: 140353; MUID:9522656

A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIM1>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
R:Lindehl, V.; Aa, K.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIM2>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAA47429.1; PID:g39777
A:Experimental source: strain CR-2
R:Wolf, M.; Geczi, A.; Boriss, R.
submitted to the EMBL Data Library, December 1993
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in *Bacillus subtilis*:
A:Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 10-508 <NOL>
A:Cross-references: EMBL:Z29076; NID:g509266; PIDN:CAA82317.1; PID:g509267
R:Seo, Y.S.; Lee, Y.H.; Pek, U.H.; Kang, H.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:44:15 ; Search time 131.22 Seconds
(without alignments)
73.041 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGFVSTLTLYDANGNP.....TIVNGPYGLRETSRLSTVFT 300

Scoring table:
BIOSOM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues
Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------|--------------------|
| 1 | 692 | 42.5 | 383 | 1 MANA_STRLI | P51529 streptomyc |
| 2 | 601 | 36.9 | 1331 | 1 MANB_CALSA | P22533 caldocellum |
| 3 | 202 | 12.4 | 409 | 1 GUN2_BACSA | P06565 bacillus sp |
| 4 | 201 | 12.3 | 400 | 1 GUN5_BACAG | O85465 bacillus ag |
| 5 | 200 | 12.3 | 488 | 1 GUN1_BACSA | P05566 bacillus sp |
| 6 | 184.5 | 11.3 | 505 | 1 GUNV_BACSA | O47096 erwinia car |
| 7 | 180.5 | 11.1 | 444 | 1 GUNN_ERMCA | O53394 erwinia car |
| 8 | 180 | 11.0 | 448 | 1 GUNN_ERMCA | P15704 clostridium |
| 9 | 177.5 | 10.9 | 466 | 1 GUN5_THERFU | O01786 thermomom |
| 10 | 174.5 | 10.7 | 504 | 1 GUNN_ERMCA | O55395 erwinia car |
| 11 | 173.5 | 10.6 | 499 | 1 GUN1_BACSU | P07983 bacillus su |
| 12 | 162.5 | 10.0 | 499 | 1 GUN2_BACSU | P10475 bacillus su |
| 13 | 161 | 9.9 | 459 | 1 GUN1_BACSU | P27035 streptomyc |
| 14 | 152 | 9.3 | 800 | 1 GUN1_BACSU | P06564 bacillus sp |
| 15 | 151.5 | 9.3 | 499 | 1 GUN3_BACSA | P23549 bacillus sp |
| 16 | 151.5 | 9.3 | 825 | 1 GUN3_BACSA | P18570 bacillus sp |
| 17 | 149.5 | 9.2 | 941 | 1 GUN_BACSA | P14924 bacillus sp |
| 18 | 147.5 | 9.0 | 429 | 1 GUN1_BACSA | P23541 butyrivibri |
| 19 | 142 | 8.7 | 562 | 1 GUN1_BACSA | P25583 acidothermu |
| 20 | 141 | 8.7 | 312 | 1 GUN4_RUMAL | O07940 ruminooccu |
| 21 | 131 | 8.0 | 814 | 1 GUN4_RUMAL | P10477 clostridium |
| 22 | 130 | 8.0 | 426 | 1 GUN_BACSU | P17974 butyrivibri |
| 23 | 122.5 | 7.5 | 1039 | 1 GUNB_CALSA | P10474 c endogluc |
| 24 | 122 | 7.5 | 438 | 1 EXG_CALSA | P29177 candida alb |
| 25 | 117 | 7.2 | 426 | 1 GUN2_ERMCH | P07103 erwinia car |
| 26 | 115.5 | 7.1 | 747 | 1 GUN2_CELFI | P50400 cellulomona |
| 27 | 113.5 | 7.0 | 517 | 1 GUN4_CLOLO | P54937 clostridium |
| 28 | 113.5 | 6.8 | 475 | 1 GUN4_CLOLO | P17901 clostridium |
| 29 | 109.5 | 6.7 | 440 | 1 GUNB_CLOCL | P28621 clostridium |
| 30 | 105 | 6.4 | 364 | 1 GUN4_RUMAL | P23660 ruminooccu |
| 31 | 104 | 6.4 | 566 | 1 GUNG_CLOLM | O05332 clostridium |
| 32 | 100.5 | 6.2 | 748 | 1 GUNG_PSEFL | P27033 pseudomonas |
| 33 | 99.5 | 6.1 | 448 | 1 EXG1_YEAST | P23776 saccharomyc |

| | | | | | |
|----|------|-----|-----|--------------|---------------------|
| 34 | 98 | 6.0 | 584 | 1 GUN2_CLOCE | P25472 clostridium |
| 35 | 97.5 | 6.0 | 375 | 1 GUN1_ROBSP | P23044 robillarda |
| 36 | 97.5 | 6.0 | 647 | 1 CN16_ECOLI | P08331 escherichia |
| 37 | 97 | 6.0 | 493 | 1 GUN4_XANCP | P19487 xanthomonas |
| 38 | 95 | 5.8 | 515 | 1 GUN2_CLOCL | P28623 clostridium |
| 39 | 94.5 | 5.8 | 645 | 1 LIP1_PHOU | P40601 photorhabdu |
| 40 | 94.5 | 5.8 | 856 | 1 AAP1_YEAST | P37898 saccharomyc |
| 41 | 94 | 5.8 | 562 | 1 EXG2_YEAST | P52911 saccharomyc |
| 42 | 94 | 5.8 | 890 | 1 BCN5_CLOPE | P08696 clostridium |
| 43 | 93.5 | 5.7 | 406 | 1 GUN1_RUMAL | P16216 ruminooccu |
| 44 | 93.5 | 5.7 | 510 | 1 YDPI_SCHPO | O13692 schizosach |
| 45 | 93.5 | 5.7 | 566 | 1 GUNB_PAEIA | P23550 paenibacilli |

ALIGNMENTS

```

RESULT 1
MANA_STRLI
ID MANA_STRLI STANDARD; PRT: 383 AA.
AC P51529;
DT 01-OCT-1996 (Rel. 34, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DR 30-MAY-2000 (Rel. 39, Last annotation update)
DE MANNAN ENDO-1,4-BETA-MANNOSIDASE PRECURSOR (EC 3.2.1.78) (BETA-
MANNAOSE) (1,4-BETA-D-MANNAN MANNANOHYDROLASE).
GN MANA.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-42.
RC STRAIN-66 / 1326;
RA MEDLINE: 93207541.
RX Arcand N., Klupefel D., Paradis F.W., Morosoli R., Shareck F.;
RT "Beta-mannanase of Streptomyces lividans 66: cloning and DNA sequence
RL Biochem. J. 290:657-663(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN-66 / 1326;
RA Shareck F.;
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: OPTIMAL ENZYME ACTIVITY IS OBTAINED AT 58 DEGREES
CC CELSIUS AND A PH OF 6.8.
CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-
CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
CC GALACTOGALUCOMANNANS.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M92297; AAA26710.2; -.
DR INTERPRO: IPR001547; -.
DR PFM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1
FT CHAIN 36 383 MANNAN ENDO-1,4-BETA-MANNOSIDASE.
FT DOMAIN 336 340 POLY-GLY.
FT SEQUENCE 383 AA; 39681 MW; 5DB4B407C64E94C3 CRC64;
SQ
Query Match 42.5%; Score 692; DB 1; Length 383;
Best Local Similarity 46.7%; Pred. NO. 2.9e-46;

```

| | | | | | | | | | | |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|--------------|----------|------------|------|--------|----|------|----|
| | MATCHES | 141; | Conservative | 51; | Mismatches | 104; | Indels | 6; | Gaps | 5; |
| OY | 2 | AANSCTVYSGLTLLPYDANGPFPVRKGINBHNATKKQDAITLBIANTGNTARYIYLSDGG | 61 | | | | | | | |
| Dd | 36 | AAGGHVSNGRVEEENGSAFWRGNNAHYTPDR-TSISIDIAIKANKTYAVTLSSGR | 94 | | | | | | | |
| OY | 62 | WTKDDIHFRNLISAENHLVAPEVDATGY--DSIASLRADVITYEMRSAL-IGKE | 118 | | | | | | | |
| Dd | 95 | WTKSASBSVALIGCKAKNKVICYLEVEDHTICYGADGATSILDQAQDVWGKRSMANROE | 154 | | | | | | | |
| OY | 119 | DTVIINIANEFSGSWEGDAMADGIYAIPRLNLAAGLNHFTLANADAAGKCQ-PPOSIDHYGR | 177 | | | | | | | |
| Dd | 155 | DVVVNININEPGNNINYAAMDATSKAISGKLKGAIGBHALMADPAANNCGDMGSIGTRSNA | 214 | | | | | | | |
| OY | 178 | EVENNDPCQRNFTESTHMETRYAGKNMSOVRTINDIVLODALVIGEFGHRHGPDVENT | 237 | | | | | | | |
| Dd | 215 | SVFASDPDRNRIFYFSJHMTG-VYDTAEARDYILNALVGNGCLPTVGESEFDQSBDGPDEDA | 273 | | | | | | | |
| OY | 238 | IIMSYSEORGVGTAWLSMKRGNEWEYTLDLSNDMAIGNLTAMGNTTVNGBYGLRETSRLST | 297 | | | | | | | |
| Dd | 274 | IMATAOSLGVLGYLGSWSMGSGGVEYELDKVMNGFDPNSITSWMGNRIILYSGINGIAATSRTAT | 333 | | | | | | | |
| OY | 298 | VY 299 | | | | | | | | |
| Dd | 334 | VY 335 | | | | | | | | |
| <hr/> | | | | | | | | | | |
| RESULT | 2 | | | | | | | | | |
| ID | MANB_CALSA | STANDARD: | PRT: | 1331 AA. | | | | | | |
| AC | P22533; | | | | | | | | | |
| DT | 01-AUG-1991 (Rel. 19, Created) | | | | | | | | | |
| DT | 01-JUL-1993 (Rel. 26, Last sequence update) | | | | | | | | | |
| DT | 15-JUL-1999 (Rel. 38, Last annotation update) | | | | | | | | | |
| DE | BETA-MANNANASE/ENDOGUCANASE A PRECURSOR [INCLUDES: MANNAN ENDO-1,4-BETA-MANNOSIDASE A (EC 3.2.1.78) (BETA-MANNANASE) (ENDO-1,4-MANNANASE); ENDO-1,4-BETA-GUCCANASE (EC 3.2.1.4) (CELLULOSE)]. | | | | | | | | | |
| DN | MANA, | | | | | | | | | |
| OC | Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus). | | | | | | | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; Thermoanaerobacter group; Caldicellulosistruptor. [1] | | | | | | | | | |
| RX | SEQUENCE FROM N.A. MEDLINE: 93119139. | | | | | | | | | |
| RA | Glids B.D., Seul D.J., Luth E., Bergquist P.L.; The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme."? | | | | | | | | | |
| RT | Appl. Environ. Microbiol. 58:3864-3867(1992). [2] | | | | | | | | | |
| RP | SEQUENCE OF 1-346 FROM N.A. MEDLINE: 91247819. | | | | | | | | | |
| RX | Luehl E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.; Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'." | | | | | | | | | |
| RL | Appl. Environ. Microbiol. 57:694-700(1991). | | | | | | | | | |
| CC | - FUNCTION: DEBRATION OF HEMI CELLULOSES. THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGUCANASE ACTIVITIES. | | | | | | | | | |
| CC | - CATALYTIC ACTIVITY: RANDOM HYDROLAYS OF 1,4-BETA-D-MANNOSIDIC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND GALACTOGLUCOMANNANS. | | | | | | | | | |
| CC | - CATALYTIC ACTIVITY: ENDOPOLYSTS OF 1,4-BETA-D-GLUCOSIDIC LINKAGE IN CELLULOSE. | | | | | | | | | |
| CC | - MISCELLANEOURS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES CELSIUS. | | | | | | | | | |
| CC | - SIMILIARTY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES). | | | | | | | | | |
| CC | - SIMILIARTY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HDROLASES). | | | | | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpostation | | | | | | | | | |

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).

| | | |
|----|--------------------------------------------------------|--------------------------------------------|
| CC | | |
| DR | EMBL; L01257; AA871887.1; -; | |
| DR | EMBL; M36063; AA872861.1; -; | |
| DR | PIR; B43745; B43745. | |
| DR | PIR; A48954; A48954. | |
| DR | HSSP; Q06851; INBC. | |
| DR | INTERPRO; IPR001547; -; | |
| DR | INTERPRO; IPR001956; -; | |
| DR | PFAM; PF00942; CBD_3; 2. | |
| DR | PFAM; PF00150; cellulase; 1. | |
| DR | PROSITE; PS00659; GLYCOSYL_HYDROL_F5_1 | |
| KW | Hydrolase; Glycosidase; Cellulose degradation; Signal; | |
| RN | | |
| FW | Multifunctional enzyme. | |
| FT | SIGNAL | 1 41 |
| FT | CHAIN | 42 1331 |
| FT | DOMAIN | 42 325 |
| FT | DOMAIN | 326 361 |
| FT | DOMAIN | 362 518 |
| FT | DOMAIN | 519 564 |
| FT | DOMAIN | 565 720 |
| FT | DOMAIN | 721 780 |
| FT | DOMAIN | 781 1331 |
| FT | ACT_SITE | 162 162 |
| FT | ACT_SITE | 257 257 |
| FT | CONFLICT | 338 338 |
| FT | CONFLICT | 340 346 |
| SQ | SEQUENCE | 1331 AA; 146892 MW; FFBDA51BBDBDFE0 CRC64; |

| | | |
|----|--------------------------------------------------------------------|----------------------------------------------------------------|
| | Query Match | 36.9%; Score 601; DB 1; Length 1331; |
| | Best Local Similarity | 44.5%; Pred. No. 1,4e-38; |
| | Matches 125; Conservative | 42; Mismatches 108; Indels 6; Gaps 4; |
| Qy | 25 | GINHGANYKDOATTALGIANGTANTYRIVISDGGQMTKDDITHVRNIIISLAEDNHLYA 84 |
| | | : : : : : : : : : : : : : : : |
| Dd | 55 | GINHACHWRJDLDTLALGIRSMGNNSRVVSYNSGYRMTKIPASEANIIISRSIGFKA 114 |
| | | : : : : : : : : : : : : : : : : : : : : : : : : : |
| Qy | 85 | -VPFHATGY---DSIASINARVYVTETMRSLGKEPTVIINIINEMFGSGMEGDAMD 140 |
| | | : : : : : : : : : : : : : : : : : : : : : : : : : |
| Dd | 115 | IILEVHDITGEGDAACSLAOAYEWEIKSVLDGNEDEFTIINTGNEYGNNTQMWVN 174 |
| | | : : : : : : : : : : : : : : : : : : : : : : : : : |
| Qy | 141 | GYSQAIPRLRNGLNHTLTAVDAWGCO-FPOSIDHYGEVENADPOKRTMFSIHMYEYAG 199 |
| | | : : : : : : : : : : : : : |
| Dd | 175 | DTKNAIKALRDGFKHTIMVDAPNMKGODMSNMRONAGSIMADPLRNIVFSIHNYG-VY 233 |
| | | : : : : : : : : : : : : : : : : : : : : : : : : : |
| Qy | 200 | GNASOVRTNIDVLNODIALVIGEEGHRTNGDVDEATIMSYSSEORGWGLAMSKGNCP 259 |
| | | : : : : : : : : : : : : : |
| Dd | 234 | NTHAKVEEYIKSFVKDGLEPLVIGEEGHQTDDBDEALVRAKYOKIGLFESWSCGNS 293 |
| | | : : : : : : : : : : : : : : : : : : : : : : : : : |
| Qy | 260 | EWEYEDLSIDWAGNNLTANGNTIVAGPYGLRETSLSITYFT 300 |
| | | : : : : : : : : : : : : |
| Dd | 294 | YVGYLDMNVNMDPNNPTEMGQYKTNAIGTSSTPPTSTVT 334 |
| | | : : : : : : : : : : : : : : : : : : : : : : : : : |
| | RESULT 3 | |
| ID | GUN2_BACS4 | STANDARD; PRT; 409 AA. |
| AC | P06565; | |
| DD | 01-JAN-1988 (Rel. 06, Created) | |
| DT | 01-JAN-1988 (Rel. 06, Last sequence update) | |
| DT | 15-JUL-1999 (Rel. 38, Last annotation update) | |
| DN | ENDOGUCANASE B (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE). | |
| GC | CELB. | |
| OC | Bacillus sp. (strain N-4). | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | |
| OC | Bacillus/Staphylococcus group; Bacillus. | |
| RN | [1] | |
| | SEQUENCE FROM N.A. | |

```

RX MEDLINE: 87056924.
RA Fukumori F., Sashihara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RT Bacillus sp. strain N-4 and their strong homology."
RL J. Bacteriol. 168:479-485(1986).
CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M14729; AAA22299.1; -.
DR PIR: B25156; B25156.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase.
FT ACT_SITE 165 165 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 254 254 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 409 AA; 45690 MW; 98866066B4DA3F CRC64;

Query Match 12.4%; Score 202; DB 1; Length 409;
Best local Similarity 25.0%; Pred. No. 1.6e-08;
Matches 70; Conservative 46; Mismatches 98; Indels 64; Gaps 14;

QY 8 VSGTLLDANGNPFVNGI-NHGAWKQDQATTAEGTANTGANTVRLVLSGGQWTKDD 66
DB 40 ISNGELVNRGEPOVLKGMSSHGLQWTC-----QVNVESKMKWLDD 81
QY 67 -----IHT-----VANLSLAEDNHLVAVPEVHDATGDSIASLNR 103
DB 82 WGITVFRAMTSSGGYIEDPSYKEKVAEVAIDLIYIIDHILSDNDPNITKEA 141
QY 104 VDWYIEKRSALFKEDVYIINIANEFSGS--WEGD--AMADGYKQALPRLNAGLNHTL 158
DB 142 KDFEDEM-SELGYDPNVITYIETANEPNSDVTWNOIPIYAE--EVLIVIRANDPNNTI 197
QY 159 MVDAAGNGQFQSIHDYGREVFADPQRTMFSIHMEYAGNAGNSOVRTNIDRYLNDPLA 218
DB 198 IV---GRTGMSQDVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLADQVDYALDQGA 249
QY 219 LVIGEFGRHRTNGD---VDEATL-MSYSEORGVTGLAWS 253
DB 250 IFVSEGTSEATGDGVFLDEAQWIDFMDERNLISMANWS 289

RESULT 4
GUN5_BACAG STANDARD; PRT; 400 AA.
ID GUN5_BACAG
AC 085465;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENDOLUCANASE 5A (EC 3.2.1.4) (EMBO-1,4-BETA-GLUCANASE) (ALKALINE
DE CELLULASE).
GN CEL5A.
OS Bacillus agaradhaerens (Bacillus agaradhaerens).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 8721;
RA Bjornvad M.E.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[2]

```

```

RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
RC STRAIN-AC13 / NCIMB 40482;
RX MEDLINE: 98153671.
RA Davies G.J., Dauter M., Brzozowski A.M., Bjornvad M.E.,
RA Andersen K.V., Schueler M.;
RT "Structure of the Bacillus agaradhaerens family 5 endoglucanase at 1.6-
RT A and its cellobiose complex at 2.0-A resolution."
RL Biochemistry 37:1926-1932(1998).
RX [3] X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
RP STRAIN-AC13 / NCIMB 40482;
RC MEDLINE: 98384136.
RA Davies G.J., MacKenzie L.F., Varrot A., Dauter M., Brzozowski A.M.,
RA Schueler M., Withers S.G.;
RT "Snapshots along an enzymatic reaction coordinate: analysis of a
RT retaining beta-glucoside hydrolase."
RL Biochemistry 37:1107-1173(1998).
CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF067428; AAC19169.1; -.
DR PDB: 1A3H; 16-MAR-99.
DR PDB: 2A3H; 16-MAR-99.
DR PDB: 3A3H; 16-MAR-99.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT_SITE 165 165 PROTON DONOR.
FT ACT_SITE 254 254 NUCLEOPHILE.
SQ SEQUENCE 400 AA; 44702 MW; 3F9C66FB9BC36FF CRC64;

Query Match 12.3%; Score 201; DB 1; Length 400;
Best local Similarity 26.6%; Pred. No. 1.8e-08;
Matches 72; Conservative 44; Mismatches 109; Indels 46; Gaps 14;

QY 8 VSGTLLDANGNPFVNGI-NHGAWY-----KQATTAEGTANTGANTVRI 54
DB 40 ISNGELVNRGEPOVLKGMSSHGLQWYGFVNVESKMKLND-----DWGINVFER 89
QY 55 VL---SDGGWTKDGI-HTVRLISLAEDNHLVAVPEVHDATGYDSIASLNAVYWIEMR 111
DB 90 AMYTSSSGGYIDDPYKVKYKAVPAEVAIDLIYIIDHILSDNDPNITKEAKPFDEM- 148
QY 112 SALIGKEDVYIINIANEFSGWEGDAMADGYK---QALPRLNAGLNHTLMDVDAAGNQ 167
DB 149 SELGDPNVITYIETANEPNS--DVTWNOIKPYAEVITPIIRNDPNNTIIV---GRT 203
QY 168 FPGSIHDYGREVFADPQRTMFSIHMEYAGNAGNSOVRTNIDRYLNDLALVIGEGFGR 227
DB 204 WSDQVH-HAADNQLADP--NVMYAFHF--YAGTHGQNLADQVDYALDQGAIFVSEGT 258
QY 228 HINGD---VDEATL-MSYSEORGVTGLAWS 253
DB 259 AATDGGVFLDEAQWIDFMDERNLISMANWS 289

RESULT 5
GUN1_BACS4 STANDARD; PRT; 488 AA.
ID GUN1_BACS4
AC P06566;

```

```

DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE A (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
GN CELA.
OS Bacillus sp. (strain N-4).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87056924.
RA Fukumori F., Sasahara N., Kudo T., Horikoshi K.;
RT "Nucleotide sequences of two cellulase genes from alkalophilic
RL Bacillus sp. strain N-4 and their strong homology.";
RL J. Bacteriol. 168:479-483(1986).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14781; AAA2301.1; -.
DR PIR; A25156; A25156.
DR INTERPRO: IPR001547; -.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR CELLULOSE degradation; Hydrolase; Glycosidase.
FW ACT_SITE 163 163 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 252 252 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 488 AA; 54264 MW; 97248B453D28D3B5 CRC64;

Query Match 12.3%; Score 200; DB 1; Length 488;
Best Local Similarity 24.6%; Pred. No. 2.8e-08;
Matches 70; Conservative 49; Mismatches 103; Indels 62; Gaps 14;

OY 3 NSGFVSTGLVDANGNPFVNRGI-NHGHWKQDQATTAIEGANTGANTVRIYVSDGQ 61
DB 33 NGQSLQNGQLVNEHGDVQLKMSHGLQWTG-----QFVNTDSIK 74
OY 62 WTKDD-----IHT-----VRNLISLAEDNHLVAPEVDATGYDSTA 98
DB 75 WLRDWMGITVFRAAMYTSSGYIEDPSVKEKVAEALDGLIYIIDMHLSDNPRI 134
OY 99 SLNRADVWIERKSLICKEDVYIINANWFG---SWEGDAMADGY-KQALPRLNAGL 154
DB 135 YKEEKKEFDEE-SALYGYIPVNYIIEINEPNGHNVR--DSHIRPYAEVTPVIRAMP 191
OY 155 NHTLWDAAGQGFQPSIDHYGREYFNADPQRTMFSIMYEYAGNAGNSQVNTINDRYLN 214
DB 132 NNIVIVGRTWS---QDVHEADNOLD-DP--NMYATFHF-YAGHGGQLKNQVDYALS 243
OY 215 ODALVIGFEGHRTNGD---VDEATI-MSYSEORGVGWGLMS 253
DB 244 RQALIFVSEMTSATGQGVLEDEAQAQWIDEMDERNLISWAMNS 287

RESULT 6
GUNV_ERMCA STANDARD: PRI; 505 AA.
AC 047096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE V PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V)
DE (CELLULOSE V).

```

```

GN CELV.
OS Erwilia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SCRI93; -.
RX MEDLINE; 94067016.
RA Cooper V.J.C., Salmund G.P.C.;
RT "Molecular analysis of the major cellulase (celv) of Erwilia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -1- FUNCTION: ENDOGLUCANASE WITH SOME EXOGLUCANASE ACTIVITY. THE PH
CC OPTIMUM IS ABOUT 7.0 AND THE TEMPERATURE OPTIMUM ABOUT 42 DEGREES
CC CELSIUS.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76000; CAA53592.1; -.
DR HSSP; Q06851; INBC.
DR INTERPRO: IPR001547; -.
DR INTERPRO: IPR001956; -.
DR PFAM; PF00942; CBD_3; 1.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR CELLULOSE degradation; Hydrolase; Glycosidase; Signal.
FW SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 505 ENDOGLUCANASE V.
FT DOMAIN 32 334 CATALYTIC.
FT DOMAIN 335 352 LINKER.
FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
FW ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 505 AA; 54900 MW; DBEA937BBAD2623 CRC64;

Query Match 11.3%; Score 184.5; DB 1; Length 505;
Best Local Similarity 23.1%; Pred. No. 4.6e-07;
Matches 71; Conservative 46; Mismatches 123; Indels 67; Gaps 12;

OY 13 LYDANGNPFVNRGI-NHGHWKQDQATTAIEGANTGANTVRIYVSDGQVTKDD----- 66
DB 48 LVDEQGRVQRLRGISSHGLQWFGD-----YNNKSMKWLNRDWMGINV 89
OY 67 -----IHTVRNLISLAEDNHLVAPEVDATGYDSTASLNRADVDTI 108
DB 90 FRVAVTTAADGYISNPSLANVKEVAQAQSLGYIIIDMHLSDNPRIYKAAKTFEA 149
OY 109 EMRSLICKEDVYIINANWFG---SWEGD--ANADGYKQALPRLNAGNHTLWDAAG 164
DB 150 EM-AGLYGSPNVYIIEINEPNGVTVNGQLRPYA---LEVTDIRSKDPNLTIV--G 202
OY 165 WGFQPSIDHYGREYFNADPQRTMFSIMYEYAGNAGNSQVNTINDRYLNODLALVIGEF 224
DB 203 TGTWSQDIDHD---AADNLPDPNTWYALHF---YAGTHQQLFRDRDYAQSAGALFVSEW 257
OY 225 GHRHTNGD-----VDEATIMSSEORGVGWGLMSKNGPMEWYL---DLSNDWAGNLL 275
DB 258 GTSDSAGSGGPFLEPSQWIDFNNRGVSWYNNWLSITDSEASALAPASNSGGWTEGNTL 317
OY 276 TAWGNTI 282

```

```

RESULT      7
ID           GUNN_ERMCA          STANDARD;          PRT;         444 AA.
AC           Q59394;
DT           01-NOV-1997 (Rel. 35, last sequence update)
DT           01-NOV-1997 (Rel. 35, last annotation update)
DE           ENDOGALACTANASE N PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-D-GALACTANASE N)
GN           CELN.
OS           Erwinia carotovora.
OC           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC           Pectobacterium.
RN           [1]
RP           SEQUENCE FROM N.A.
RX           STRAIN-ATROSEPTICA FCBR C18;
RX           MEDLINE: 98299944.
RA           Olsen O., Thomsen K.K., Weber J., Dues J.O., Svendsen I., Wegener C.,
              von Wettstein D.;
RT           "Transplanting two unique beta-galactanase catalytic activities into
              one multienzyme, which forms glucose.";
RL           Biotechnology 14:71-76(1996).
CC           -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
              LINKAGES IN CELLULOSE.
CC           -1- SUBCELLULAR LOCATION: SECRETED.
CC           -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
              HYDROLASES).
-----
CC           THIS SWISS-PROT entry is copyright. It is produced through a collaboration
              between the Swiss Institute of Bioinformatics and the EMBL Outstation -
              the European Bioinformatics Institute. There are no restrictions on its
              use by non-profit institutions as long as its content is in no way
              modified and this statement is not removed. Usage by and for commercial
              entities requires a license agreement (see http://www.isdb.ch/announce/
              or send an email to license@isdb.ch).
-----
EMBL: L39788; AAC37033.1; -.
DR           INTERPRO: IPR001547; -.
DR           INTERPRO: IPR001956; -.
DR           PFAM: PF00942; CBD_3; 1.
DR           Pfam: PF00150; cellulase; 1.
DR           PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KM           Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT           SIGNAL                      1               31
FT           CHAIN                       32             444
FT           ACT_SITE                     168            168
FT           ACT_SITE                     256            256
SQ           SEQUENCE   444 AA;  48300 MW;  FA7E4179004CB843 CRC64;
              NUCLEOPHILE (BY SIMILARITY).
Query Match      11.1%; Score 180.5; DB 1; Length 444;
Best Local Similarity 22.5%; Pred. No. 7.9e-07;
Matches 69; Conservative 48; Mismatches 123; Indels 67; Gaps 12;

13 LYDANGNFVWRCI-NRHHAWYKQQTATAIEBIANTGANTVAIVYSDDGGQTKND----- 66
Db 48 LVDEGGKRHVQLKGVSHTGHWFGD-----TYKKDSMKRWLRDMGWINV 89
Oy 67 -----IHVTRNTLSIAEDNHVLVAPEVDHATGYDSTIASLNRAVDWI 108
Db 90 FRVAVMTAADGVISNPISLANRYKEAVALAAQSIGYIIIDMHLISNDPRTIYAQAKTFFA 149
Oy 109 EKRSALLIKEDTVITINIANEF--SWBGD-AWADYKQAIPLRLNAGLNTLTAVDAAG 164
Db 150 EM-AELYGSSPNVIYEINEIRNGSGVTWNGQIRPYA---LEVDTIRSKDPDLITIV--G 202
Oy 165 WGFQFSITHDYGREVFENDPQRENTESIMHEVEAGNNSQVETNTIDRYLVNDLALVIGE 224
Db 203 TGTWQODTHD---AADNQLPDPTNIYALHF-YAGTHGOFLDRIDVAYSRGCAALFVSEW 257

```

```

0Y 225 GRRHNGG-----VEAATIMSYSEORGWNTLMSKNGNPEMEYI-----DLSDNAGNLT 275
Db 258 GTSDAKSGNGGPPPLPSQWIDPFLNKGYSWVWMSLTDKSEASALAPAKSKGGRTQNL 317
QY 276 TANGNTI 262
Db 318 STSGKEV 324

RESULT 8
GUN_CLOAB STANDARD: PRT: 448 AA.
ID GUN_CLOAB
PI5704:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN EGLA.
OS Clostridium acetobutylicum
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PP262;
RX MEDLINE; 88268074.
RA Zappe H., Jones W.A., Jones D.T., Woods D.R.;
RT "Structure of an endo-beta-1,4-glucanase gene from Clostridium
RT acetobutylicum p262 showing homology with endoglucanase genes from
RT Bacillus spp.";
RL Appl. Environ. Microbiol. 54:1289-1292(1988).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION OF C.ACETOBUTYLICUM IS NOT
CC REQUIRED FOR ACTIVITY.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce,
CC or send an email to license@sib-sib.ch).
CC -----
CC CC EMBL; M3311; AAA23230.1; -.
CC DR PIR; A27631; A27631.
CC DR INTERPRO; IPR001547; -.
CC DR PFAM; PF00150; cellulase; 1.
CC DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5_1
CC KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC FT SIGNAL 1 34
CC FT CHAIN 35 448
CC FT ACT_SITE 175 175 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 263 263 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT_SITE 263 263
CC SQ SEQUENCE 448 AA; 49366 MW; 6EB6AC97C836A598 CRC64.

Query Match 11.08; Score 180; DB 1; Length 448;
Best Local Similarity 22.18; Pred. No. 8.7e-07;
Matches 71; Conservative 59; Mismatches 139; Indels 56; Gaps 14;

QY 8 VSGTLLYANANPNPYMGR-I-NHGAAWY-----KQATTAIESIANTGANTYAI 54
Db 49 VVGSQLDCSNKSPKLPQIKLMSHGLQWYNAFYVDSMKELRDK-----WGVNVIRA 98
QY 55 VL--SDGQGW-I-KDITHTVNLISLADNNHVAPEVHADAGYDSIASLNRAVYWIIM 110
Db 99 AMYINNEGYSINPSSQSKRIKIKIYQADALDLNNKTYIIDHHILSDNNPNNTYIKQAKSFFQEM 158
QY 111 RSALIGKEDTVIINIANEMFGSWEGDANAWGK---QAIPLRLNAGNHTLIMVDAAGWG 166

```



```

Db 159 AEE-YGKSYNVIYEICNPNNG---GTNNANDIKPYANTIIIPAIIRIDPNNTIIIVGISTWS 214
Oy 167 QEPQSIHDYGRVFNADPQR--NTWFSIHTMEYAGNAGNAGVNTIDRYLNDLALVIGEE 224
Db 215 QDVYDIAAD-----NPLFRYSNIMYTCHE--YAGTHQSLRDJINVMKGIAPVTEW 264
Oy 225 GHRHTNGD----VDEA-TIMSYSEORGVGNLMSKNGKNGPEMEYL---DLSDNMAGNNTL 275
Db 265 GTSDASGNGGPFIDESQKWPDMASKNISMWNALCDKSEASALKSGSTTGCTGWTSD 324
Oy 276 TANGNTIYNGPYGLRSTRLS 296
Db 325 TTSGLEFVKKSIGSGNTTSGTS 345

RESULT 9
GUNS_THREU STANDARD; PRT; 466 AA.
ID GUNS_THREU STANDARD; PRT; 466 AA.
AC 001786:
DT 01-JUL-1993 (Rel. 36, Created)
DT 15-JUL-1993 (Rel. 38, Last sequence update)
DT 15-JUL-1993 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE E-5 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE E-4)
DE (CELLULOSE E-5) (CELLULOSE E5).
GN CELE.
OS Thermomonospora fusca.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Nocardiopsaceae; Thermobifida.
RN RP SEQUENCE FROM N.A.
RX STRAIN-YX;
RX MEDLINE; 91258320.
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT "DNA sequences of three beta-1,4-endoglucanase genes from
RT Thermomonospora fusca";
RT J. Bacteriol. 173:3397-3407(1991).
RN RL [2]
RN REVISIONS.
RA Lao G., Ghangas G.S., Jung E.D., Wilson D.B.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN RP [3]
RN SEQUENCE OF 137-142 AND 157-166.
RC STRAIN-YX;
RA Irwin D.C., Spezio M., Walker L.P., Wilson D.B.;
RT "Activity studies of eight purified cellulases: specificity,
RT synergism, and binding domain effects.";
RT Biotechnol. Bioeng. 42:1002-1013(1993).
RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
CC -1- PATHWAY: CELLULOSE DEGRADATION.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01577; AAC09379.1; -
CC PIR; C42360; C42360.
CC HSSP; P07986; IEXH.
CC INTERPRO; IPR001547; -
CC INTERPRO; IPR001919; -
CC PFM: PF00553; CBD_2; 1.
CC PFM: PF00150; cellulase; 1.
CC PROSITE; PS00659; GLYCOSYL-HYDROL_5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC SIGNAL
CC CHAIN 1 36 ENDOGLUCANASE E-5.
CC FT 37 466

```

```

FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 391 391 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 466 AA; 49800 MW; 1CF0ADFBE2DEF82E CRC64;

Query Match 10 9%; Score 177.5; DB 1; Length 466;
Best Local Similarity 24.8%; Pred. No. 1.4e-06;
Matches 78; Conservative 49; Mismatches 121; Indels 67; Gaps 18;

Oy 8 VSGTLYDANGNPFVARGIN-HGHANYKDOAT-TALEGIA-NTGANTVRI---VLSGQ 61
Oy 173 VCGTQLDEHGNPVQKRGSTHGIOWFCHLSDSLALAYPMKADIIILSMYIEDGYE 232
Db 62 WT-----KDIATVRLTSLAEDNHLVAVPEVHATGQYDLSIANGVWIERSLICK 117
Db 233 TNRGFTDRMH---QLIDNATARGLYIVDMHTLTGDPHYINDRDKTKTFAELAQHASK 289
Oy 118 EDPIITIANENFGSWEGDAD--GY-KOALPRLNAGLHTLNVDAAGQFPQSIHD 174
Db 290 TN-VLTETANE-----PQGVMSIKSYAEVYIPVIRQDPDSVIYVTRGWSLGVY-EG 343
Oy 175 YGREVENADP--QRTWFSIHTMEYAGNAGVNTIDRYLNDLAL---VIGFGRHT 229
Db 344 SGPAELIANPVMASNIMYAFHFY-----AASHRDYLNALREASLEFPVYTFETET 397
Oy 230 NGD-----VDEATISYSEORGVGNLMSKNGKNGPEMEYLIS 267
Db 398 TGDANDPQMDR--YIDLAERKIGTKWNSDDPFRSGAVPQRCASGP----- 447
Oy 268 NDWAGNNTLWANGNTI 282
Db 448 --WSGSLKASGQWV 460

RESULT 10
GUNS_ERWCA STANDARD; PRT; 504 AA.
ID GUNS_ERWCA STANDARD; PRT; 504 AA.
AC 059395:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ENDOGLUCANASE V1 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V1)
DE (CELLULOSE V1).
GN CELV1.
OS Erwina carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
RN RP [1]
RN SEQUENCE FROM N.A.
RX STRAIN-SCC3193;
RX MEDLINE; 95231512.
RA Mae A., Helkinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwina carotovora subspecies
RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in
RT phytopathogenicity";
RT Mol. Genet. 247:17-26(1995).
RL -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X79241; CAAS5823.1; -
CC HSSP; O06851; INBC.
CC INTERPRO; IPR001547; -

```

```

DR INTERPRO: IPR001956; -.
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31
FT CHAIN 32 504
FT DOMAIN 32 334
FT DOMAIN 335 352
FT DOMAIN 353 504
FT ACT_SITE 168
FT ACT_SITE 256
FT ACT_SITE 256
SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match
Best Local Similarity 10.7%; Score 174.5; DB 1; Length 504;
Matches 73; Conservative 43; Mismatches 123; Indels 69; Gaps 14;

13 LYDANGNPPVPMGT-NHGMATKQDATTABEGTANTGANTVRIYVSDGQTKDD-IHT 69
48 LYDEGKRVQLRGISSNGIQWGD-----YVNRKDSKMTLRDDMGIV 89
70 VNLISLAEENHLVAVPEY-----HDANGYDSIASLNRADV 107
90 FVAVATTAE-NGYIANPSLANKYKEAVAAAGLGYITIIDHTLSNDNYTKAOKTF 148
108 IEMRSALIGKEDTVIINIANENFGS--SWEGD--AMADGYKQAIPLRNAGLNTLWADA 163
149 AAM-AGLVNSPNVIEIANEPNGSVTNGQIRPYA---LEVTDIRSDPNLTIYVSG 204
164 GNGQPPQSHDYGREYFNADPQATWESIMTEYAGNAGNSQVNTINDRLNDLALVIE 223
205 TWS---ODIHD---AADNLPDPNTLYALHF--YAGTHQGFDRIDYQSGAALFVSE 256
224 FGHRTNGD-----VDEATINSYSEQGVGLAMSKNGKPEMEYL---DISNDAGNN 274
257 WTSIDASGGGFFLPESQTWIDPLNNRGLISVWMSLSDKSETSALVAGASGGWTEON 316
275 LTWAGNTI 282
317 LSTSGKEV 324

RESULT 11
GUN1_BACSU STANDARD; PRT; 499 AA.
AC P07983;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN BGIC OR GID.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-DIG;
RC MEDLINE: 87194581.
RA Robson L.M., Chambliss G.H.;
RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DIG.";
RL J. Bacteriol. 169:2017-2025(1987).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16185; AAA22496.1; ALT_INIT.
DR PIR: A26874; A26874.
DR HSSP: 006851; INRC.
DR INTERPRO: IPR001547; -.
DR INTERPRO: IPR001956; -.
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169
FT ACT_SITE 257 257
FT ACT_SITE 350 499
FT DOMAIN 350 499
SEQUENCE 499 AA; 55187 MW; 339D04E95A63EE1 CRC64;

Query Match
Best Local Similarity 10.6%; Score 173.5; DB 1; Length 499;
Matches 69; Conservative 68; Mismatches 122; Indels 73; Gaps 15;

3 NGFVYSGTLYDANGNPPVPMGT-NHGMATKQDATTABEGTANTGANTVRIYVSDGQ 61
39 NQLSIKGTGLVNRDQKAVQLGKISSHGLQWYGD-----FVNRKDSIK 80
62 WTKDD-----IHT-----VRLISLAEENHLVAVPEYVHATGYDSA 98
81 WLRDMGIVFPAATYATDGGTIDNPVKNKYKEAVEAKELGIVYIIDHRLINDGNPQ 140
99 SLNRADVIEKRSALIGKEDTVIINIANENFGS--SWEGD--AMADGYKQAIPLRNAGL 154
141 MEKAKKEPKEM-SSLYGTPVIVIEIANEPNGDVNMKRDIPYAE--EVSIVRKNDP 196
155 NHTLWADAAGWQFPQSHDYGREYFNADPQ--RTWFSIMTEYAGNAGNSQVNTINDR 211
197 DWIITV---GTGWSQDVND-----AADQLKDAENVYALHF--YAGTHQGLDRKAN 245
212 VLNODLALVIGFGRHRTNGD-----VDEA-TIMSYSEQGVGLAMSKNGKPEMEYL-- 264
246 ALSKAPLFTVETWGTSDASNGGVLPDQSRWLANLIDSKNISVWMSLSDKSESSALKP 305
265 --DISNDAGNNLTWAGNTIYVNGPYGLRSTR 294
306 GASKTGWPLTDLTASGTFVRENIRGTIDSTR 337

RESULT 12
GUN2_BACSU STANDARD; PRT; 499 AA.
AC P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCSE) (CELLULOSE).
GN BGIC OR GID OR BELS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-PAP15;
RC MEDLINE: 8706793.
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-CK-2;
RX MEDLINE: 95225655.
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
RL subtilis CK-2."
RN Antione Van Leeuwenhoek 66:327-332(1994).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Rose M., Entian K.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RN SEQUENCE OF 30-45.
RC STRAIN-CK-2;
RX MEDLINE: 95225655.
RA Aa K., Flensburg R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
RL endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
RN compost soil."
RN Antione Van Leeuwenhoek 66:319-326(1994).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)."
CC -----
DR EMBL: Z29076; CA82317.1; -
DR EMBL: X04689; CA28392.1; -
DR EMBL: X67044; CA47429.1; -
DR EMBL: Z73234; CA97610.1; ALT_INIT.
DR EMBL: Z99113; CAB13696.1; ALT_INIT.
DR PIR: A26114; A26114.
DR HSP: 006851; INBC.
DR SUBTILIST; BG10A37; BGIC.
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR001956; -
DR PFAM: PF00942; CBD_3; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 29
FT CHAIN 30 499
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EAEZ CRC64;

```

```

Query Match 10.0%; Score 162.5; DB 1; Length 499;
Best Local Similarity 21.0%; Pred. No. 2.2e-05;
Matches 66; Conservative 75; Mismatches 136; Indels 37; Gaps 15;

```

```

QY 3 NSGFYVSGTLYDANGNPFVARGI--NHGAWYKQDQT-TAIGANTGANTV--RIYLS 57
DB 39 NGOLSTKIGQLVNRDCKAVQKISHSGLQWGEVYKMSLKLMDQDGIYFRAMATYA 98
QY 58 DGGQWTKRDI-HTVNLISLAEDNLVAVPEVDATGYSIASLNRANDYTIEMSAALIG 116
DB 99 DGGYIDNPVSKYKKEAVEAKELGIYIILDMHILNDGNPNONKEKAEFFREM--SSLYG 157
QY 117 KETVLIINANWFG--SWBGD--ANADGYKAIPLRLNAGLHRLTAVDAAGCOFPOSI 172
DB 158 NTDNVIYELANPEKNGVDNKRDKIPYAE--EYISYIKRNDPNIIIV--GIGTMSQOV 211
QY 173 HDYGREVFNDPQ--RNTWFSIHMYETAGNAGNSQVRTNIDRLVNDLALVIGEGHRT 229

```

```

DB 212 ND-----AADDOLKDNAMVALHF--YAGTGOFLRKANALSKGAPITVTEGTSDA 263
QY 230 NGD-----VDEA-TIMSYSHORGWGLAMSKNGCEMEYL-----DLSNDMAGNNITANGN 280
DB 264 SGNGYFLDQSRWELKYLDKSKITISWNWNLSDKQSSSALKRGASKTGWRSLDSASGT 323
QY 281 TIVNGPYGLRSTSR 294
DB 324 FVRENILGTRDSTK 337

```

```

RESULT 13
GUNA_STRLI STANDARD; PRT; 459 AA.
ID GUNA_STRLI
AC P27035;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE ENDOGLUCANASE CELA PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULOSE).
GN CELA.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-43.
RC STRAIN-66;
RX MEDLINE: 92246492.
RA Theberge M., Lacaze P., Shareck F., Morosoli R., Kluepfel D.;
RT "Purification and characterization of an endoglucanase from
RT Streptomyces lividans 66 and DNA sequence of the gene."
RL Appl. Environ. Microbiol. 58:815-820(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -----
CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)."
CC -----
DR EMBL: M82807; -; NOT_ANNOTATED_CDS.
DR HSP: P07986; IEXH.
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR001919; -
DR PFAM: PF00553; CBD_2; 1.
DR PFAM: PF00150; cellulase; 1.
DR PROSITE: PS00651; CBD_BACTERIAL; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 27
FT CHAIN 28 459
FT DOMAIN 136 135 ENDOGLUCANASE CELA.
FT DOMAIN 147 147 CELLULOSE-BINDING.
FT DOMAIN 148 357 LINKER ("HINGE") (PRO-THR BOX).
FT ACT_SITE 286 286 CATALYTIC.
FT ACT_SITE 378 378 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 31 131 BY SIMILARITY.
SQ SEQUENCE 459 AA; 48663 MW; CFE47EC34E2F97A7 CRC64;

```

```

Query Match 9.9%; Score 161; DB 1; Length 459;
Best Local Similarity 25.0%; Pred. No. 2.6e-05;
Matches 81; Conservative 42; Mismatches 145; Indels 56; Gaps 16;

```

```

SQ SEQUENCE 800 AA; 88602 MM; 7CCAAD7B6ADAD55CF CRC64;

Query Match 9.38; Score 152; DB 1; Length 800;
Best Local Similarity 22.88; Pred. No. 0.00026;
Matches 74; Conservative 53; Mismatches 125; Indels 72; Gaps 18;

QY 12 TLVDANGPFYWRGYN-HGHAWYKQDAI-TALEGANT-GANTVATLV--SDGQWTKDD 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 67 TLVDQHGRIKIDRGKSTHGLQFPILINDNAYKALANWESNMIRLAVYENGASNP 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 IHTVNL--ISLAEDNEHVAPEVY-DATGYDSIASLRADVWYIEMKSLIGKEDYII 123
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 127 LKRSVYIGIDLATENDMTVYVDMEVAAPDPRDPYTAGAEDEFEDL--AALYPPNPHTLY 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 NIANEMFSGWEGDA-----WADGYKQALP--RLNAGL--NHTLVDAAGWGCFPQ 170
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 ELANPSSNNNGAGGAI-PNNEGMAVVKRYAPIYEMLRDSCNADNIIITVSPNNQRPD 245
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 171 SIHDYGREYVFNADPQRTWFSIHMT-----ETAGGNAQRYNRINDRYLNDL 217
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 LAAD-----NPIDHHTMYTVHFTGSHAASTESYPPETPNSERGNVMSNTRYALENG 299
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 ALVIGEGEHRHRTNGD---VDEATI--MYSQGRGWTLAN-----WKG 256
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 300 AVFATQWTSQANGGSGPYFDEADWIEFLNENNISMANWSLTJNKREVSQAFTPPELGRS 359
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 N-----GPE--W--EYLDLSNDW 270
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 360 NATSIDPSPDQWVPEELSLSGEY 383

RESULT 15
GUN3_BACSU
ID GUN3_BACSU STANDARD; PRT; 499 AA.
AC P23549;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE ENOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCASE) (CELLULOSE).
GN BGIC.
OS Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
| 1 |
RC SEQUENCE FROM N.A.
RC STRAIN-BSE616;
RX MEDLINE; 91299280.
RX PARK S.H., Kim H.K., Pack M.Y.;
RT "Characterization and structure of the cellulase gene of Bacillus
RT subtilis BSE616.";
RL Agric. Biol. Chem. 55:441-448(1991)
-1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
LINKAGES IN CELLULOSE.
-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announcements/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01057; BAA00859.1; -
DR PIR; J0111; J0111.
DR HSSP; 006851; INBC.
DR INTERPRO; IPR001547; -
DR INTERPRO; IPR001956; -
DR PFAM; PF00942; CBD_3; 1.
DR PFAM; PF00150; cellulase; 1.

```

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
FM Cellulose degradation: Hydrolase; Glycosidase; signal.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 499 AA; 55169 MW; 2E821E3D8BAC404 CRC64;

Query Match 9.38; Score 151.5; DB 1; length 499;

Best Local Similarity 20.78; Pred. No. 0.00016;
Matches 65; Conservative 75; Mismatches 137; Indels 37; Gaps 15;

```
OY 3 NSGFVSGTLYDANGNPFVARGI-NHGHAWYKDOAT-TAIEGIANTGANTV---RIVLS 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 39 NGQLSIKGTQLYNRDGRAVOQLKGISSHQLQWYGEVNNDSIKWLRDQWGLVFRAMYTA 98
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 58 DGGQWTKDDI-HTVNNLSLAEDNHLVAVPEVHDGTGSTASLNRAVDYTIEMRSALIG 116
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 99 DGGIDNPSVKKMKKEAVEAKELGIYIIDWHILNDGNPNQNKKEKAEFFKEM-SSLYG 157
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 117 KEDYIINIANEMFG--SMED--AMADYKQAIPLRNAGLNHTIATVDAAGWGQFPOSI 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 158 NTPNVIYEIANEFNGDVNMKDKIPYAE--EVLIVIRKNDPNITIV---GTGWSQDV 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 173 HDYGREVFNADPO--RNTMFSIHMYEYAGNAGSVRTNIDRYLNQDLALVIGFGRHT 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 212 ND-----AADDQLKDAVMDALHF--YAGTHGQFLRDKANYALSKGAPIFVTEWGTSDA 263
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 230 NCD----VDEA-TIMSYSQKRGVGLAWSKRGNGPEWEYL---DLSNDWAGNNLTAGN 280
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 264 SGNQGVFFLDQSRKWLKYLDSKTSISVNMNLSDKQSSSALKFQASKTGQWRLSDLSASGT 323
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 281 TIVNGPYGLRSTR 294
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 324 FVRENILGTRKDK 337
```

Search completed: December 19, 2000, 16:44:17
Job time: 893 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 19, 2000, 16:17:09 ; Search time 363.46 Seconds
(without alignments)
77.071 Million cell updates/sec

Title: US-09-339-159-2_COPY_31_330
Perfect score: 1630
Sequence: 1 NANSGFYVSGTLYDANGNP.....TIVNGPYGLRHSRLSTVFT 300

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues
Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_14:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mhcc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 1172 | 71.9 | 516 | 2 | 066185 bacillus ci |
| 2 | 797 | 48.9 | 395 | 2 | 069347 vibrio sp. |
| 3 | 710 | 43.6 | 384 | 2 | 086599 streptomyce |
| 4 | 634 | 38.9 | 279 | 2 | 092F13 thermomonos |
| 5 | 614 | 37.7 | 930 | 2 | 09RFX5 |
| 6 | 536.5 | 32.9 | 327 | 2 | 09R115 |
| 7 | 226 | 13.9 | 1097 | 2 | 09ZAI7 |
| 8 | 197 | 12.1 | 557 | 2 | 09ZAI7 anaerobic t |
| 9 | 191 | 11.7 | 389 | 2 | P94622 clostridium |
| 10 | 176.5 | 10.8 | 501 | 2 | 059232 bacillus sp |
| 11 | 174.5 | 10.7 | 499 | 2 | 045532 bacillus su |
| 12 | 170.5 | 10.5 | 930 | 2 | 059280 clostridium |
| 13 | 166.5 | 10.2 | 387 | 2 | 031029 erwinia car |
| 14 | 162.5 | 10.0 | 486 | 2 | 043430 bacillus sp |
| 15 | 161.5 | 9.9 | 499 | 2 | 052731 bacillus sp |
| 16 | 160.5 | 9.8 | 481 | 2 | 066064 actinomyces |
| 17 | 159.5 | 9.8 | 635 | 2 | 066065 fibrobacter |
| 18 | 157 | 9.6 | 783 | 2 | 045554 bacillus sp |
| 19 | 154.5 | 9.5 | 749 | 2 | 059154 anaerocellu |

| | | | | | |
|----|-------|-----|------|---|----------------------|
| 20 | 152 | 9.3 | 821 | 2 | 059241 bacillus sp |
| 21 | 140 | 8.6 | 278 | 2 | 09RKS5 streptomyce |
| 22 | 136.5 | 8.4 | 570 | 2 | 059665 pseudomonas |
| 23 | 129 | 7.9 | 478 | 5 | 016028 globodera r |
| 24 | 126 | 7.7 | 319 | 5 | 018454 heterodera |
| 25 | 126 | 7.7 | 438 | 3 | 09URL8 globodera alb |
| 26 | 126 | 7.7 | 476 | 5 | 018453 heterodera |
| 27 | 125 | 7.7 | 395 | 5 | 09T5M4 globodera t |
| 28 | 123 | 7.5 | 319 | 5 | 061595 heterodera |
| 29 | 123 | 7.5 | 319 | 5 | 077449 heterodera |
| 30 | 121.5 | 7.5 | 1000 | 2 | 024820 thermophil |
| 31 | 117 | 7.2 | 391 | 5 | 077094 globodera r |
| 32 | 117 | 7.2 | 392 | 5 | 044078 globodera r |
| 33 | 116.5 | 7.1 | 332 | 2 | 060054 unidentified |
| 34 | 116 | 7.1 | 426 | 2 | 09REW0 erwinia chr |
| 35 | 115 | 7.1 | 363 | 2 | 007652 celivibrio |
| 36 | 114 | 7.0 | 419 | 3 | 012539 agardicus b1 |
| 37 | 114 | 7.0 | 470 | 5 | 09U6M5 globodera t |
| 38 | 113 | 6.9 | 910 | 3 | P87211 orphomyces |
| 39 | 112.5 | 6.9 | 621 | 2 | 007653 celivibrio |
| 40 | 111.5 | 6.8 | 506 | 5 | 09UAS7 meloidogyne |
| 41 | 110 | 6.7 | 435 | 3 | 012626 picchia anqu |
| 42 | 109.5 | 6.7 | 754 | 2 | 085318 salmoneilla |
| 43 | 107 | 6.6 | 430 | 3 | 012712 trichoderna |
| 44 | 106 | 6.5 | 325 | 3 | 09Y8H6 emericeila |
| 45 | 105 | 6.4 | 456 | 3 | 09Y7Y7 schizosacch |

ALIGNMENTS

| RESULT | 1 | PRELIMINARY; | PRT; | 516 AA. |
|--------|----------------------------------------------------------------------|--------------|------|---------|
| 066185 | 066185 | | | |
| ID | 066185 | | | |
| AC | 066185 | | | |
| DT | 01-AUG-1998 (TIREMBLrel. 07, Created) | | | |
| DT | 01-AUG-1998 (TIREMBLrel. 07, Last sequence update) | | | |
| DT | 01-MAR-2000 (TIREMBLrel. 13, Last annotation update) | | | |
| DE | MANNAVASE. | | | |
| OS | Bacillus circulans. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | |
| OC | Bacillus/staphylococcus group; Bacillus. | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE; 98233274. | | | |
| RA | Yoshida S., Sako Y., Uchida A.; | | | |
| RT | *Cloning, sequence analysis, and expression in Escherichia coli of a | | | |
| RT | gene coding for an enzyme from Bacillus circulans K-1 that degrades | | | |
| RT | guar gum." | | | |
| RL | Biosci. Biotechnol. Biochem. 62:514-520(1998). | | | |
| DR | EMBL; AB007123; BAA25878.1; - | | | |
| DR | INTERPRO; IPR001547; - | | | |
| DR | PFAM; PF00150; cellulase; 1. | | | |
| SO | SEQUENCE 516 AA; 55245 MW; D8E5A48642AE8062 CRC64; | | | |

Query Match 71.9%; Score 1172; DB 2; Length 516;
Best Local Similarity 69.2%; Pred. No. 2e-84;
Matches 207; Conservative 50; Mismatches 42; Indels 0; Gaps 0;

| | | |
|----|-----|------------------------------------------------------------------|
| QY | 1 | NANSGFYVSGTLYDANGNPVFMGINHGRAMVDQATTAEGTANTGANTVRIYLSGG 60 |
| DB | 33 | HAASGFYVSGTLYDANGNPVFMGINHGRAMVDQATTAEGTANTGANTVRIYLANGH 92 |
| QY | 61 | QWTEDDIHTVNNLSLAEDNHLVAVPEVHDATGYDSIASINRAVDYTEMASATIGREDT 120 |
| DB | 93 | KWIDDVNTVNNILTLEQNKLLAVLEVDHATGSDSLDNNAVYWGKSLATIGREDR 152 |
| QY | 121 | VVINIANEWFSGWEDAMADYKQAIPIRLNAGINHTLWYDAAGWGFOPSHDYGREVF 180 |
| DB | 153 | VVINIANEWFSGWEDAMADYKQAIPIRLNAGINHTLWYDAAGWGFOPSHDYGREVF 212 |
| QY | 181 | NADPQRTMFSIMHYETAGAGNASGVRTNIDRYLNMQDLALVIGFGRHRTNGDVDEATIMS 240 |

[illegible]

```

RESULT 2
069347
ID 069347; PRELIMINARY; PRT; 395 AA.
AC 069347;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-1,4-MANNANASE.
GN MANA.
OS Vibrio sp.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NA-138;
RA Tamau Y., Araki T., Morishita T., Sakita K., Ohmura K.;
RL J. Ferment. Bioeng. 83:201-205(1997).
EMBL: D86329; BAA25188.1; -
DR INTERPRO: IPR001547; -
DR INTERPRO: IPR002883; -
DR PFAM: PF00150; cellulase; 1.
DR PFAM: PF02013; CBD_5; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN 1.
SQ SEQUENCE 395 AA; 43098 MW; A3E30B3A978C5DA1 CRR64;

```

| | | | | |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match | 48.98; | Score 797; | DB 2; | Length 395; |
| Best Local Similarity | 50.88; | Pred. No. 4.4e-55; | | |
| Matches 155; Conservative | 55; | Mismatches 87; | Indels 8; | Gaps 5; |

| | | | | |
|----|--|-----|----------------------------------------------------------------|-----|
| OY | | 2 | ANSEFYSGTLTXANGPEFMRGIRINGHAWYKQAOATLEGINTANCANTVRYLSPGGQ | 61 |
| Dd | | 19 | AHAFFYSNGVLYANGSBAFKIRGINAHETWTDKLVSALSGIATAPANTVRYVALSNGR | 78 |
| OY | | 62 | WTCKDITVRNLISLAEDNHLVAPEVDHATGY--DSIASLRNAVYWIEMRSALIGKE | 118 |
| Dd | | 79 | WTKNDVSDVYNIIIMLAANNIALLEHDTTGCEBESSASLSDAADYWEIKNELIGQE | 138 |
| OY | | 119 | DTVLIINIANEF--SWEGDAWADYKOAIPLRLNAGLNHTLMWDAAWGQ-FPOSIDYG | 176 |
| Dd | | 139 | DYVLILINIGNEPFGNNDIVAVMVDHVSAIQRLBSAGINHTIMVDAPMMGDGMKFMLNNA | 198 |
| OY | | 177 | REVENADPQRKTMSIHAYE-YAGGNASOVNTINDRVLNQDLALYTEFGHRRTNGCYDE | 235 |
| Dd | | 199 | QEVNSDPKLTITISVHYEITYSSYNS--VNDYISSFTNNGLVYLIGEFASTHERGAUYDE | 256 |
| OY | | 236 | ATIKSYSRQRCVGMLAWSMGNGSEWELDLSNDMAGNNULTAQONTIYNNGPYGIAREDSRL | 295 |
| Dd | | 257 | GSIERSBSTLSLGiyIGWSGMSGNDPTTSDLDIVNMMDNNSYSTGWNVLINONGOSIKSTSL | 316 |
| OY | | 296 | STEFT 300 | |
| Dd | | 317 | ATVEFT 321 | |

| RESULT | 3 | | |
|--------|---------------------------------------------------------|------------------------------------------|--------------|
| 086599 | | | |
| ID | 086599 | PRELIMINARY; | PRT; 384 AA. |
| AC | 086599 | | |
| DT | 01-NOV-1998 | (TIREMBLrel. 08, Created) | |
| DT | 01-NOV-1998 | (TIREMBLrel. 08, Last sequence update) | |
| DT | 01-MAY-2000 | (TIREMBLrel. 13, Last annotation update) | |
| DE | SECRETED BETA-MANNOSIDASE. | | |
| GN | MANNA2. | | |
| OS | Streptomyces coelicolor. | | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | |

CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2).
 RA Oliver K., Harris D.,
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2).
 RA Parkhill J., Barrell B.G., Rastandream M.A.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2).
 RX MEDLINE: 97000351
 RA Redenbach M., Kleiser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Kinash J.H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL031514; CAA20610.1; -
 DR INTERPRO: IPR001547; -
 DR INTERPRO: IPR002883; -
 DR PfAM: PF00150; cellulase, 1.
 DR PfAM: PF02013; CBD_5, 1.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 SQ SEQUENCE 384 AA; 39847 MW; B8AD158B3FD3BD11 CRC64;

| | | | | |
|-----------------------|------------------|------------------|-----------|-------------|
| Query Match | 43.68; | Score 710; | DB 2; | Length 384; |
| Best Local Similarity | 46.7%; | Pred. No. 3e-48; | | |
| Matches 141; | Conservative 53; | Mismatches 102; | Indels 6; | Gaps 4. |

[illegible]

| | RESULT | |
|--------|----------------------------------------------------------------------|--------------|
| 09ZF13 | 4 | |
| 09ZF13 | | PRELIMINARY; |
| 09ZF13 | | PRT; 279 AA. |
| AC | 01-MAY-1999 (TREMBLrel. 10, Created) | |
| DT | 01-MAY-1999 (TREMBLrel. 10, Last sequence update) | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last annotation update) | |
| DE | BEHA-MANNANASE (EC 3.2.1.76) (FRAGMENT). | |
| GN | MAN. | |
| OS | Thermomonospora fusca. | |
| OC | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | |
| CC | Actinomycetales; Streptosporangiales; Nocardiopsaceae; Thermobifida. | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN-KW3; | |

RA Hlilge M., Gloor S.M., Winterhalter K., Zimmermann W., Piontek K.;
 RT "Crystallisation and preliminary crystallographic analysis of two
 RT beta-mannanase isoforms from Thermomonospora fusca KM3.";
 RL Acta Crystallogr. D 52:1224-1225(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-KW3;
 RC MEDLINE: 99036862.
 RA Hlilge M., Gloor S.M., Rypniewski W., Sauer O., Reichtman T.D.,
 RA Zimmermann W., Winterhalter K., Piontek K.;
 RT "High-resolution native and complex structures of thermostable beta-
 RT mannanase from Thermomonospora fusca - substrate specificity in
 RT glycosyl hydrolase family 5.";
 RL Structure 6:1433-1444(1998).
 CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-BETA-D-MANNOSIDIC
 CC LINKAGES IN MANNANS, GALACTOMANNANS, GLUCOMANNANS, AND
 CC GALACTOGLUCOMANNANS.
 CC EMBL: AJ006227; CA006924.1; -.
 DR INTERPRO: IPR001547; -.
 DR PFAM: PF00150; cellulase; 1.
 DR PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; UNKNOWN_1.
 KW Hydrolyase; Glycosidase.
 FT NON_TER 1
 FT 279 279
 SQ SEQUENCE 279 AA; 30657 MW; 98A2850FAE013DB0 CRC64;

Query Match 38.9%; Score 634; DB 2; Length 279;
 Best Local Similarity 44.4%; Pred. No. 1.8e-42;
 Matches 128; Conservative 56; Mismatches 82; Indels 22; Gaps 6;

QY 5 GFVSGTLYDANGNPVPMGNGHAYKDOQTALTEGANTGATVRYVSDGGQWTK 64
 DB 1 GLTVKNGRYLEANGQEEIINGVSHPHNMY-POHTQAPADIKSGANTVRVLSNGVWMSK 59
 QY 65 DDHITVRNLSLAEDNHLVAPVEVDATGY---DSIASLRADYVTEKMSALIGKEDTV 121
 DB 60 NGSDVANVSLCKONRFLICMLTEVHDTGTGEGSGASTLDQADVTELSVYQGEEDTV 119
 QY 122 IINIANEMF-----GSWEGDAMADYKQALPRLNRNGLHNTLVADAAGNQ-EPQS 171
 DB 120 LINGNEPEYNDSATVAAGADTSA-----AIQRLRAAGFEHTLVADPNNGQDWTNT 172
 QY 172 IHDYGEVFADPQRTMFSIHMYEYAGNAGSOVRNIDVNLQDLALVYGECHRTNG 231
 DB 173 MRNNADQVVASDPTGNTVFISIMYG-VYSGASTITSTLEHFNAGPLIIGEGHDSG 231
 QY 232 DVDDEATIMSYEORGVGLAMSWKNGPEWEYIDLSDMAGNLTLAG 279
 DB 232 NPDEDTIMAEERLKIGYIGMSNGSGNGGVEYLDVNTNFDGDLSPWG 279

RESULT 5
 ID 09RFX5 PRELIMINARY; PRT; 930 AA.
 AC 09RFX5;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
 DE MULTIDOMAIN BETA-1,4-MANNANASE PRECURSOR.
 GN MANA.
 OS Caldicellus cellulosorans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Caldicellus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sunna M.D., Bergquist P.L.;
 RT "A novel multidomain beta-1,4-mannanase gene from Caldicellus
 RT cellulosorans and action of the recombinant enzyme on krat pulp.";
 RL EMBL: AF163837; AAF22274.1; -.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.

DR INTERPRO: IPR002965; -.
 DR PFAM: PF00150; cellulase; 1.
 DR PFAM: PF00942; CBD_3; 2.
 DR PRINTS: PRO1217; PRICEXTENS.
 DR PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; UNKNOWN_1.
 KW SIGNAL.
 FT SIGNAL 1
 FT CHAIN 34 33 POTENTIAL.
 FT SEQUENCE 930 AA; 101516 MW; 0086538D54D1A2CC CRC64;

Query Match 37.7%; Score 614; DB 2; Length 930;
 Best Local Similarity 46.9%; Pred. No. 3.9e-40;
 Matches 122; Conservative 52; Mismatches 80; Indels 6; Gaps 5;

QY 25 GINHGAWKRDQATALTEGANTGATVRYVSDGGQWTKDDHITVRNLSLAED-NHLY 83
 DB 486 GINHPHAWYDRLLSSLOGIRSGANAVRYVLSNGCRKIPASEVADITISQARTIGYRA 545
 QY 84 AVEVDATGY--DSTA-SINRAVDYVTEKMSALIGKEDTVIINIANEMFGSWEGDAMAD 140
 DB 546 VVLEVDHTGTGEGDAACSTTAVNTEKNTLAGQENFVYINIGNEPEYNNYQWYV 605
 QY 141 GYKQALPRLNAGLNTLWADAAGNQ-EPQSIHDYGEVFADPQRTMFSIHMYEYAG 199
 DB 606 DTRNAVOALNNAGINNTINWDAPNMQDSFTREDNAPTFINADPORNLFVSIHMYG-VY 664
 QY 200 GNASOVRTNIDRYLQDLALVIGFGRHNGDVEATIMSYEORGVGLAMSWKNGP 259
 DB 665 DTAEEVQSYIESYVNGGLPLVIGEPHMSDGPNDQALVQYAKQINIGLFGMSNGSG 724
 QY 260 EWEYLDLSDNAGNNTFANG 279
 DB 725 GVEYLDVNTNFDGDLSPWG 744

RESULT 6
 ID 09RUT5 PRELIMINARY; PRT; 327 AA.
 AC 09RUT5;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
 DE PUTATIVE SECRETED BETA-MANNOSIDASE (FRAGMENT).
 GN MANA.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denaparte D., Elchner A., Cullum J.,
 RA Kinash H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AF121746; CAB57406.1; -.
 DR INTERPRO: IPR001547; -.
 DR PFAM: PF00150; cellulase; 1.
 FT NON_TER 327
 FT SEQUENCE 327 AA; 34772 MW; A21B0068C2C0975D CRC64;

Query Match 32.9%; Score 536.5; DB 2; Length 327;
Best Local Similarity 41.9%; Pred. No. 1.1e-34;
Matches 114; Conservative 52; Mismatches 97; Indels 9; Gaps 7;

QY 4 SGGVSTTLYDANGNPFVARGIN-HGH-AWKDQATTAEGIANATGA-NTVRIYVSDG--- 59
DB 57 AGHIGDGLLENGNDNFVARGINHAHTWY-PGE-TOSIADYKALGNANSVAVYVLSDBHRS 115
QY 64 KDDHRTVRLISLAEDNHLVAVPEVHDATGY--DSIA-SLNRAVDYIEMRSALIKEDPT 120
DB 116 ENGADVAVAIEOCKANRLICVLEHDTGTGAEDAAAGTIDHADYVIGIKDVLACQEDV 175
QY 121 VIINIANEFWSGEGAMADGYKQALPRKAGLNTITMDAAGKQCFPSI-HDGRFY 179
DB 176 VIVNIGNEPWNTPDPACTEPTVAAYKRLRAAGLQHTIMDAPRWGDQWYRANRSY 235
QY 180 FNDPQNTFSEIMYEXAGNSQVRTNIDRVNLQDALVIGHF-GHRTNGDVDEAT 238
DB 236 YNDPQNTLIFSIMYS-VEDTQAEITDYANAFDALPILIGFGPADQYGDPEEDTM 294
QY 239 MYSERQGVGLWMSWKGNGPEWEYIDLSDW 270
DB 295 MATAEQRLRIGYIAMSWSGNTD--PVLIDLADF 324

RESULT 7
Q92A17 PRELIMINARY; PRT; 1097 AA.
AC Q92A17;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE S-LAYER ASSOCIATED MULTIDOMAIN ENDOGLUCANASE.
GN CELA.
OS anaerobic thermophile KM-THCJ.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KM-THCJ;
RX MEDLINE; 99175466.
RA Cann I.K., Kocherginskaya S., King M.R., White B.A., Mackie R.I.;
RT "Molecular cloning, sequencing, and expression of a novel multidomain
mannanase gene from Thermococcus bacterium polysaccharolyticum";
RL J. Bacteriol. 181:1643-1651(1999).
DR EMBL; U82253; AAD09354.1; -.
DR INTERPRO; IPR001119; -.
DR INTERPRO; IPR001547; -.
DR INTERPRO; IPR002088; -.
DR PFAM; PF00150; cellulase; 1.
DR PFAM; PF00395; SLH; 3.
DR PFAM; PF02018; CBD_6; 2.
DR PROSITE; PS00904; PPTA; UNKNOWN_1.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
DR SEQUENCE 1097 AA; 119762 MW; C41C6B92C0494B9 CRC64;

Query Match 13.9%; Score 226; DB 2; Length 1097;
Best Local Similarity 24.8%; Pred. No. 1.7e-09;
Matches 83; Conservative 60; Mismatches 134; Indels 58; Gaps 15;

QY 6 FVVSSTLYDANGNPFVARGIN-HGH-AWKDQATTAEGIANATGA-NTVRIYVSDG--- 59
DB 41 FHVAVNKIYDPGNDPFIKGVNIGYRSWEKSVLQDVHLVDMKFTVRLNCTIGNN 100
QY 60 ---GWTYKDDITVRLNISLAEDNHLVAVPEVHDATGDSIA-----SLNRAVDY 107
DB 101 WEGGTGANNIDDAI---IKAFTRKRYVVEIDLDHTGYPLSNPPAPGQPSLDQAI-AM 156
QY 108 IEMRALKCKEDTYIINIANEFWSME--GDAMADGYKQALPRKAGLNTITMDAAG 165
DB 157 FELAKKIDNPFVWNTNNEPSSAPLDPQWKYANEITIAINSTGADNIYVD--GW 214
QY 166 GQFPOSIDH-----YGREVFNADPQRTWFSIMHYEXAGNSQVRTNIDR 211

DB 215 SYANGIGQNPPTVEKRSVAVLYTGQDGLNDSKANTIFAFHNTN-EGDIQKVEDYDR 273
QY 212 VINODLALVIGFEGHRTNGVDENTIMSYE---QREVGGLWSM-----KGN 257
DB 274 ANAKGLVYFMEYEGYDYS--DAAKEGVRSGLQAVNKGAGRIYWNWDGYLDYLTSGTGR 331
QY 258 GEWEYIDLSDMAGNNLTAMGNTIVNGPYGLRET 292
DB 332 GSGWE-TNKTDGSKPTLNSWGDKIWDNNGITPT 365

RESULT 8
P94622 PRELIMINARY; PRT; 557 AA.
AC P94622;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ENDO-1,4-BETA GLUCANASE ENGF (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
DE (ENDO-1,4-BETA-GLUCANASE) (CARBOXYMETHYL CELLULOSE).
GN ENGF.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97136706.
RA Shewetta S.A., Ichl-Isht A., Park J.S., Liu C., Malburg L.M.,
RA Doi R.H.;
RT "Characterization of enfF, a gene for a non-cellulosomal Clostridium
RT cellulovorans endoglucanase.";
RL Gene 182:163-167(1996).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
DR EMBL; U37056; AAB40891.1; -.
DR HSSP; 085465; 2A3H.
DR INTERPRO; IPR001547; -.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSTYL_HYDROL_F5; UNKNOWN_1.
KW Hydrolyase; Glycosidase.
SQ SEQUENCE 557 AA; 60131 MW; D186EC8EB504EED CRC64;

Query Match 12.1%; Score 197; DB 2; Length 557;
Best Local Similarity 26.7%; Pred. No. 1.2e-07;
Matches 88; Conservative 33; Mismatches 132; Indels 76; Gaps 17;

QY 12 TLYDANGNPFVARGIN-HGHAWTKDQATTAEGIANATGA-----NTVRIYV--SD 58
DB 62 TLCDKDNPIQLNGMSTHGLQMF-----PGVNNNAFFALSDMNSNVIRLAMYAE 113
QY 59 GGGWTKDDI-HYVRNLISLAEDNHLVAVPEVHDATGDSIASLNRAVDYIEMRSALIGK 117
DB 114 GGYATNPSYKQTYINGINVAINDMYIVDMHMANPQDNASVYSGAQSFNDISTLYER 173
QY 118 EDVYIINIANEFWSGEG---DA--WADGYKQALP---RLNAGLNTITMDAAGQEP 169
DB 174 NKNITTYLCEPENGENGNGVATNDATGMAQVSYATPIQLRNKGNELITVGNPFSQBR 233
QY 170 QSHIDYREYFNADPQRTWFSIMHYEXAGN-ASQVRYTIDR-----VINODLALY 221
DB 234 DLAMD-----NPINDNTIYSVHF--YSGTNPISVDTRDNAMSNVRYALNHAAVA 295
QY 222 GEGHRTNGDVDEATIMSYSEORGWGLWMSWKGNGPEWEYIDLSDN----- 269
DB 286 TEGWTSIATG-----TGPTLV-AKADAMLDF-LNGNNISMCSNISIKDKKAALNSLTS 338
QY 270 -----WAGNNLTAMGNTI---VNGPY 287
DB 339 LDPGSKLMDNDELTTSGQYVRAIRKAY 367


```

DB 39 NGOLSKTQVLYNKGKAVOLKGISSHGLQWYGD-----FVKNKSLK 80
OY 62 WTKD-----IHT-----VRNLISLAEDNHLVAVPEVDATGYSIA 98
ID 81 WLRDMGJTFVFRRAATYTDGGYIDNPVSKNRKVEAKELGIVYIDMHIIDNGPNQ 140
OY 99 SLNRAVDWIERKALIGKEDVTIINIANEMEG--SWECD--ANADGYKQAIPLRNAGL 154
DB 141 HEKAKDFPEK--SLYKGTNPVYIEINERPGDYNMRDKRPYAE--EYISVYRKNDP 196
OY 155 NHTLWADAGWQFPOSIDHYREVFNADPO--RNFESITMYIAGNAGNAGVPTNDR 211
DB 197 DNIIIV---GTGWSQDVND-----AADDQLKANDVYALHF--YAGTHQSGLDKRANY 245
OY 212 VANDLALVIGFGRHNTGD-----VDEA--TIMSYSEORGWGLAWSKNGPEWEYL-- 264
DB 246 ALSKAPITVTEGTSIDAGNGGVFLDQSRWMLNIDSKNISVWNNLSDKQESSALKP 305
OY 265 --DLSNDMAGNNLTWANGTIVNGPYGLRETSR 294
DB 306 GASKTGWPLTDLTASTGTFRNLTGNKDKSTK 337

RESULT 12
O59290 PRELIMINARY; PRT; 930 AA.
ID 059290
AC 059290
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-1,4-ENDOGLUCANASE (EC 3.2.1.4).
GN CELA.
OS Clostridium josui.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN [1]
RP SEQUENCE FROM N.A.
RA Fujino T., Fujino E., Karita S., Ohmura K.;
RT "Revised sequence of cels gene encoding endoglucanase (Eg)-1 from
RT Clostridium josui."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; D85526; BAA12826.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001119; -.
DR INTERPRO; IPR001547; -.
DR PFAM; PF00150; cellulase; 1.
DR PRAM; PR00395; SLH; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Hydrolase; Glycosidase.
KW SEQUENCE 930 AA; 101726 MW; 0EBAE8754D084448 CRC64;

Query Match 10.5%; Score 170.5; DB 2; Length 930;
Best Local Similarity 26.0%; Pred. No. 3.2e-05;
Matches 76; Conservative 45; Mismatches 106; Indels 71; Gaps 19;

```

```

OY 234 DEATIMS---YSEORGWGLAWSW-----KGNP-----EM-EYLDLSN-DNAGNNLT 276
DB 265 DRGNISNARVIALEHGVAVFCSEMGTSSEASGNNGYIKLEADWLEFLIANNISWIMSLT 324

RESULT 13
O31029 PRELIMINARY; PRT; 387 AA.
ID 031029
AC 031029
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
GN CELA.
OS Erwinia carotovora subsp. carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-LY34;
RA Park Y.W., Lim S.T., Yun H.D.;
RL Mol. Cells 0:0-0(1997).
RL EMBL; AF025768; AAC02964.1; -.
DR HSSP; O85465; 2A3H.
DR INTERPRO; IPR001547; -.
DR PFAM; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Signal; Hydrolase.
KW SIGNAL 31
FT CHAIN 32 387
FT FT BETA(1,4)-GLUCAN GLUCANOHYDROLASE.
SQ SEQUENCE 387 AA; 42003 MW; 6302B3DF8AC2738B CRC64;

Query Match 10.2%; Score 166.5; DB 2; Length 387;
Best Local Similarity 21.5%; Pred. No. 1.8e-05;
Matches 68; Conservative 50; Mismatches 113; Indels 85; Gaps 13;

```

```

OY 13 LYDANGNPFVARGI-NHGHAWYKDOATTAIEGANTGATVRYLSDGGQWTKDD----- 66
DB 48 LYDEGKRVOLRGVSSHGLQWFGD-----YKKKSMKMWLPDDWGINV 89
OY 67 -----IHTVRNLISLAEDNHLVAVPEVDATGYSIALNRAVDYWI 108
DB 90 SRVANYTADGYISKPSLIANKVKAVALAOSLYIITIDMHIISDNENPTIKYBAKTFFA 149
OY 109 EKRSAIIEKEDVTIINIANEMEG--SWECD--ANADGYKQAIPLRNAGLHTLMDAAG 164
DB 150 EM-AGLYGNSPVIYIEINERNGGVTWMDGELRPYA--LEVETIRSRDPPNLIIV--G 202
OY 165 WQFPOSIDHYREVFNADPOANTMFSTHMEYIAGNAGNAGVPTNDRVNDLALVIGEF 224
DB 203 TGTWSQDIDH--AADDNLDPENLTYALHF--YAGTHQGFIRIDVYAGSAGAAIFVSEW 257
OY 225 GHRHNTGD-----VDEATIMSSEORGWGLAWSKNGPEWEYLDLS----- 267
DB 258 GTSIDAGNGGPPLPSPHWIIDEPLNRG-----GSRVWMSLTDKSEASALAPGASK 308
OY 268 -NDMAGNNLTWANGTI 282
DB 309 CGGWTEQNLSASGKTV 324

RESULT 14
O45430 PRELIMINARY; PRT; 486 AA.
ID 045430
AC 045430
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CARBOXYMETHYL-CELLULOSE) (CMCSE) (CELLULOSE).
GN CEL.
OS Bacillus sp.

```

| | |
|----|-----------------------------------------------------------------------|
| CC | Bacterifirmicutes; Bacillus/Clostridium group; |
| CC | Bacillus/Staphylococcus group; Bacillus. |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN-D04; |
| RX | MEDLINE; 96029707. |
| RA | Han S.J., Yoo Y.J., Kang H.S.; |
| RT | "Characterization of a bifunctional cellulase and its structural gene |
| RT | the cell gene of Bacillus sp. D04 has exo- and endoglucanase |
| RT | activity."; |
| RL | J. Biol. Chem. 270:26012-26019.(1995). |
| CC | -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC |
| CC | LINKAGES IN CELLULOSE (BY SIMILARITY). |
| CC | -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL |
| CC | HYDROLASES). |
| DR | EMBL; U27084; AAC43478.1; -. |
| DR | HSSP; 085465; 2A3H. |
| DR | INTERPRO; IPR001547; -. |
| DR | INTERPRO; IPR001956; -. |
| DR | PFAM; PF00150; cellulase; 1. |
| DR | PFAM; PF00942; CBD_3; 1. |
| DR | PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1. |
| RR | PRODOM; PD001947; -. 1. |
| KM | Signal. |
| FT | 1 29 POTENTIAL. |
| FT | 30 486 ENDOGLUCANASE. |
| FT | DOMAIN 350 486 ENDOGLUCANASE (BY SIMILARITY). |
| FT | ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY). |
| FT | ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY). |
| FO | SEQUENCE 486 AA; 53820 MW; BD3B65BF18D6656B CRC64; |

| | Query Match | 10.0% | Score 162.5 | DB 2 | Length 486 |
|--------|--------------------------------------------------------------|-----------------------------------------------------------------|-------------|-------------------|------------|
| | Best Local Similarity | | 21.0% | Pred. No. 5.3e-05 | |
| | Matches 66 | Conservative | 75 | Mismatches 136 | Indels 37 |
| | | | | Gaps | 15 |
| QY | 3 | NSGFVYSGTTLTDANGNPFVYNGKI-NHGHAWYKDOAT-TAIGIANTGANTY---RIYLS | 57 | | |
| Db | 39 | NGQLSIFKGTOLVNRDQKAVQLKGISHSGLQWAGEYNNKSLKMLDMDGKITFPRAMTYA | 98 | | |
| QY | 58 | DGGGTRKDDI-RTVNNLISLADNHLVAVPEVHDATGYSLSLNRAVDYWTENMSALIG | 116 | | |
| Db | 99 | DGGYIDNPSYKRRKVEAVEAAKELDIYIIDIHLLINDGNPNONKKEKAKFEFFEM--SLYG | 157 | | |
| QY | 117 | KEDYTIINIANEMFG--SWEGD--AMADGKRAIPRLRNAGLNHTLMDVADAGMGQFPOSI | 172 | | |
| Db | 158 | NTPNVITIELANPPNDGAVNKKDKIKPFAE---EVSIVIRKNDPNNTIIV--GTGWSQDV | 211 | | |
| QY | 173 | HDYGRVENADPO---RNTMFSIMHYEYAGNAGSAOVRTNIDVTLNODLALVIGCFGRHT | 229 | | |
| Db | 212 | ND-----AADDQLKDANVMTALHF--YAGTIGQGLRDGANVALSKGAPIFPTWGTSDA | 263 | | |
| QY | 230 | NGD----VDEA-TINSYSEQGVCGVLANSWKGNGPEWEYL---DISNDWAGNNITLAWCN | 280 | | |
| Db | 264 | SGNGSVFELDQSGHEWKLKYIDSKTISYMWNNLSDKQESSALKPKGASKGTGWRSLDLSASGT | 323 | | |
| QY | 281 | TIYNGPYGLREMSR | 294 | | |
| Db | 324 | FVRENITLGTKDSYK | 337 | | |
| RESULT | 15 | | | | |
| ID | 052731 | PRELIMINARY; | PRT: | 499 AA. | |
| AC | 052731 | | | | |
| DT | 01-JUN-1998 | (TREMBLerelease 06, Created) | | | |
| DT | 01-JUN-1998 | (TREMBLerelease 06, Last sequence update) | | | |
| DT | 01-JUN-2000 | (TREMBLerelease 14, Last annotation update) | | | |
| DE | ENMO-B-1,4-GUCCANASE (EC 3.2.1.4) (CELLULASE) (ENDOGUCANASE) | | | | |
| DE | (ENDO-1,4-BETA-GUCCANASE) (CARBOXYMETHYL CELLULASE). | | | | |
| GN | CENS. | | | | |
| OS | Bacillus sp. | | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; | | | | |

CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Jung K.-H., Chun Y.-C., Lee J.-C., Kim J.-H., Yoon K.-H.;
 RL Biotechnol. Lett. 18:1077-1082(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-79-23;
 RA Yoon K.-H., Jung K.H.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -I- CATALYTIC ACTIVITY: ENDONUCLEOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC EMBL; AF045482; AAC02536.1; -.
 DR HSSB; O85465; 2A3H.
 DR INTERPRO: IPR001547; -.
 DR INTERPRO: IPR001956; -.
 DR PFAM; PF00150; cellulase; 1.
 DR PFAM; PF00942; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
 DR PRODOM; PD001947; ? 1.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 499 AA; 55193 MW; 990A18786CF738 CRC64;

| | | | | | |
|----|------------------------------------------------------------------|--------|--------------------|-------|-------------------------------------|
| | Query Match | 9.9% | Score 161.5: | DB 2: | Length 499; |
| | Best Local Similarity | 21.0%: | Pred. No. 6.6e-05; | | |
| | Matches | 66; | Conservative | 74; | Mismatches 137; Indels 37; Gaps 15. |
| Oy | 3 NSGPFVSGTTLTDANGNPFVNKGI-NHGHWKPOAT-TAIGIANTGANTV---RIYLS | 57 | | | |
| Dd | 39 NGQLSKTGQGLVNRGRKXAVOLKGISHGQWGEVYNDSLKMLDDMGITLYFRAMNTA | 98 | | | |
| Oy | 58 DGGQTKDGI-HTVKNLISLAEDNHLVAPEVHADATGYDSTIASLNRAVDWIEMRSALIG | 116 | | | |
| Dd | 99 DGGYIDNPSSVKRKYEAWEAAKELGITYIIDWHINDGNPNOKREKAKEFEKKEM--SLYG | 157 | | | |
| Oy | 117 KEDTVIINIANWFG--SMEGD--AMADGQKAIPIRLNAGNLHNLMDADAOWGFPOSI | 172 | | | |
| Dd | 158 NTPNVITYELANPPNDVMKRDIKPYLE--EVISIVIRKNDPDNIIY--GTGWSDOV | 211 | | | |
| Oy | 173 HDYGREVENADEQ--RNTMFSIHAYEYAGNASQVRTIDEVLNODLAYIGEFGHRHT | 229 | | | |
| Dd | 212 ND-----AADQQLDANYMALHF-YAGTIGQYLROKANALTSKGAPIFTWTGTSIA | 263 | | | |
| Oy | 230 NGD---VDEA-TINSYSBQRGVGLAWSWKNGPWEYXL---DLSDMAQNLTJLANGN | 280 | | | |
| Dd | 264 SGNGVFVLDQSHEMLKYLDISKTIYVNNWNLSDKQESSALKPKGASGTGMRLSDLSASGT | 323 | | | |
| Oy | 281 TIVNPGYGLRETISR | 294 | | | |
| Dd | 324 FVRENILDTKDATK | 337 | | | |

Search completed: December 19, 2000, 16:17:12
Job time: 1013 sec

Wed Dec 20 15:36:58 2000

us-09-339-159-2_copy_31_330_1.rspt

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)